Systematic Review of Artificial Intelligence Techniques for Blood Glucose Predicting in Diabetes

Mohammed Alotaibi
Faculty of Computers and Information Technology
University of Tabuk, Saudi Arabia
mmalotaibi@ut.edu.sa

Osama Moh'd Alia
Faculty of Computers and Information Technology
University of Tabuk, Saudi Arabia
oalia@ut.edu.sa

Abstract: *Background/Objectives: The accurate prediction of Blood Glucose Levels (BGLs) is important for the effective management of diabetes. Artificial Intelligence (AI) has become an effective tool for forecasting BGLs and offering decision support and therapeutic interventions. In this study, existing AI techniques for predicting BGLs were reviewed systematically to identify their limitations and strengths. Methods: Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, a systematic review of relevant studies was performed. Research was identified through a detailed search of several databases including PubMed, Institute of Electrical and Electronics Engineers (IEEE) Explore, and Scholar with a focus on publications ranging from 2010 to 2025. Search terms included "AI for predicting glucose levels in diabetes" and related keywords. Non-English and irrelevant studies were excluded. Results: Fifty-seven studies met the inclusion criteria. They used AI models such as Bidirectional Long Short-Term Memory (Bi-LSTM) and Convolutional Neural Network (CNN), decision trees, and hybrid models like CNN-Gated Recurrent Unit (GRU). Many models demonstrated high accuracy, especially Bi-LSTM and CNN-GRU hybrids, but were often limited by small datasets, lack of generalizability, and low interpretability. Non-invasive sensing technologies and real-time systems using Internet of Things (IoT) and wearable devices showed significant promise but require further validation. Conclusions: AI-based glucose prediction systems have advanced significantly, with some models achieving Root Mean Squared Error (RMSE) values below 10 mg/dL in short-term forecasting. However, no single approach consistently outperforms others across all prediction horizons and patient scenarios. Future research should focus on integrating multimodal inputs, enhancing model interpretability, and validating systems using diverse, real-world datasets to ensure clinical reliability.*

Keywords: *Artificial intelligence, blood glucose prediction, deep learning, diabetes management, hypoglycemia prediction, hyperglycemia prediction, machine learning.*

Received August 31, 2025; accepted January 15, 2026
https://doi.org/10.34028/iajit/23/5/12

1. Introduction

Diabetes mellitus is a chronic disease that affects millions of people across the world. Its prevalence is increasing at a disturbing rate. Per the International Diabetes Federation, 540 million individuals were living with diabetes in 2024, a number expected to double by 2050 and reach more than 1.3 billion [39, 40]. Typically, diabetes is characterized by the inability of the body to regulate Blood Glucose Levels (BGLs) in an effective manner, which leads to serious health complications. Type 1 Diabetes (T1D) usually results from an insufficient production of insulin by the pancreas, while Type 2 Diabetes (T2D) is often related to obesity and inactivity, which causes the body to use insulin in an inefficient manner [56].

The maintenance of stable BGLs is important for people with diabetes because both high (hyperglycemia) and low (hypoglycemia) can result in health complications. For example, severe hyperglycemia can lead to diabetic ketoacidosis, while hypoglycemia can result in a coma if it is not treated. It should be noted that patients are often unaware of these problems as they develop, especially during sleep, which makes their timely detection and intervention important [70, 17].

Recent developments in Artificial Intelligence (AI) have introduced promising advancements in diabetes management, especially in the prediction of BGLs before major changes take place [28]. In fact, precise prediction models could offer early warnings of hyperglycemia or hypoglycemia, which would help patients take proactive action such as adjusting their diet or taking dosage of insulin [44]. For example, a sleeping patient at risk of hypoglycemia could be awakened in time to eat and stabilize their BGL.

Furthermore, AI techniques including Machine Learning (ML) and Deep Learning (DL) models contribute to the what is known as closed-loop systems, also referred to as Artificial Pancreas (AP) systems [1], as shown in Figure 1. These systems integrate Continuous Glucose Monitors (CGMs), which monitor glucose levels at intervals ranging from one to five

minutes; AI-based algorithms that calculate insulin needs, and insulin pumps that automatically deliver the appropriate dosage [1, 13, 41].

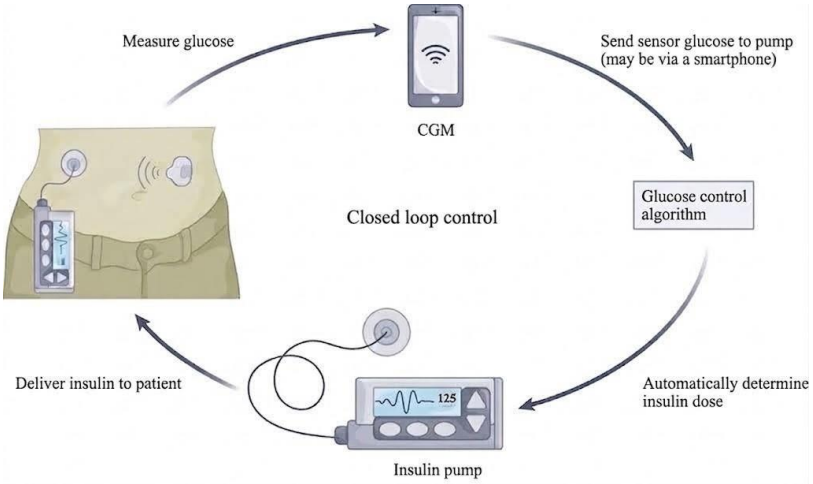

Figure 1. Closed-loop system (AP systems) [8].

In addition to predictive modeling, understanding how glucose levels are currently measured is essential. Although not the primary focus of this manuscript, it is worth briefly introducing the technologies used to monitor glucose concentrations, as these provide the foundational data required by AI-based prediction systems. Commercial technologies include traditional finger-prick glucometers [89], which rely on glucose oxidase-based reactions to measure blood glucose from a small sample as shown in Figure 2, and CGM systems, such as the Dexcom and FreeStyle Libre as seen in Figure 3, which use subcutaneous sensors to provide real-time glucose readings at frequent intervals [38]. Minimally invasive wearable patches, including microneedle and transdermal systems, also offer continuous monitoring with improved user comfort.

Figure 2. Finger-prick glucometers [Wikipedia].

On the experimental front, non-invasive sensors are being explored to enhance ease of use and reduce discomfort. These include microwave-based sensors, which infer glucose levels through dielectric changes in tissues [87]; Quartz Crystal Microbalance (QCM) and resonant sensors, which detect mass or dielectric shifts from glucose interactions on functionalized surfaces [67]; and optical approaches such as Raman and near-infrared spectroscopy [85]. Although promising, these methods often require further refinement in terms of accuracy, calibration, and real-world reliability [37, 51]. This review contextualizes AI prediction models within sensing technologies, including CGM and novel non-invasive devices, thereby enhancing the comprehension of how physiological measures support data-driven diabetes care.

Figure 3. Dexcom G7 and freestyle libre 3 systems [tcoyd.org].

This paper examines AI's contribution to predicting BGLs through a systematic review of current research. Weighing the pros and cons of current techniques will pave the way for the birth of stronger and more secure tools for the control of diabetes. Several studies are concentrating on this connection, for instance, Ahmed *et al*. [1] who presented a paper on 12 AI-based wearable-device studies spanning the years 2015-2023; Hameed *et al*. [37] who examined the recent use of ML techniques for diabetes prediction from 2018 to 2024; and Yismaw *et al*. [92] who investigated the utilization of AI in determining which T2D patients are at risk of non-adherence to treatment. Nonetheless, these studies deal with the general issue of diabetes care rather than focusing on BGL prediction. Likewise, Liu *et al*. [52] reported on machine learning models in general from 2010 to 2022 and Contreras and Vehi [19] mentioned an eight-year study from 2010-2018 without a specific focus on BGL prediction.

This review, unlike the recent surveys, introduces several separate yet important aspects to BGL prediction. First, by thoroughly reviewing 57 studies

published from 2010 to 2025 that particularly focus on BGL prediction via AI techniques, it is the most extensive and directed analysis done so far. Second, the review positions machine learning, deep learning, ensemble, and hybrid algorithms side by side, taking into account their predictive performance, methodological differences, and the suitability of the different prediction settings for each of them. Third, a very thorough study of the datasets applied in those studies is presented, including their characteristics, classification by type (real or simulated), and access. Forth, this paper provides a classification of the evaluation metrics employed in the literature, like Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R^2, and accuracy, displaying inconsistencies and pinpointing the most informative metrics for predicting BGL. Fifth, it points out the main research gaps and obstacles, such as inadequate dataset diversity, no external validation and low generalizability of the existing models, and gives easy-to-follow guidelines for improving AI-based BGL prediction.

The rest of this paper is organized as follows: Section 2 describes the materials and methods, and Section 3 presents the results and discussions. Section 4 provides a review of the datasets. Section 5 discuss the finding of the paper. The final section covers the conclusion and future directions.

2. Methods

A qualitative design was used in this study, following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure a transparent and detailed review [50]. A detailed search was carried out across several databases ranging from Scholar to PubMed and Institute of Electrical and Electronics Engineers (IEEE) Xplore to find different relevant studies. Search terms included "AI for predicting glucose levels in diabetes" and related keywords; studies published from 2010 through 2025 were considered in order to review recent developments in this field. The systematic review process involved identifying and selecting studies that met the inclusion criteria outlined in Section 2.1. The exclusion criteria are described separately in Section 2.2.

2.1. Inclusion Criteria

- Studies published in English.
- Studies focused on AI applications for the prediction of BGLs in diabetes.
- Research studies published between 2010 and 2025.

2.2. Exclusion Criteria

- Non-English studies.
- Observational research.
- Studies on AI not related to the prediction of BGLs.
- Articles that did not have sufficient methodology or data for analysis.

2.3. PRISMA Flowchart

Please see Figure 4 for the PRSIMA flowchart.

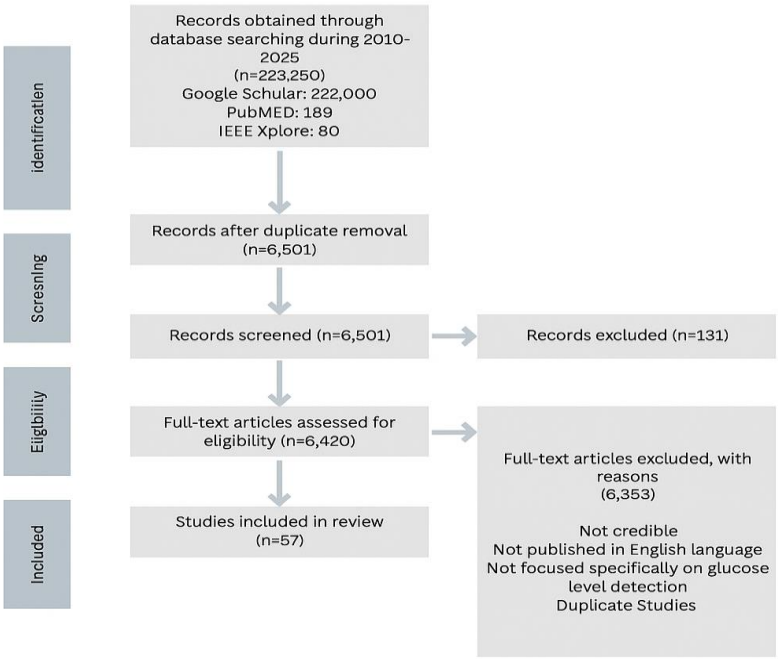

Figure 4. PRISMA flowchart.

2.4. Ethics statement

This research is a systematic review and does not involve any animal or human participants, or any other direct interventions. No ethical approval was required, but ethical principles were upheld by ensuring that all the cited studies were credited properly [26]. The findings of this review are based on data that is available publicly

from peer-reviewed journals.

3. Results

The systematic review identified 57 studies evaluating different AI techniques used to predict BGLs in people with diabetes. This systematic review compares different AI-based approaches for predicting BGLs in diabetic patients and identifies both strengths and limitations across studies. The selection of methods varied from ML models to DL, and ensemble or hybrid algorithms. For the sake of clarity and organization, ML models in Section 3.1 shall be the first to be summarized, then in Section 3.2, DL methods will be covered, and lastly in Section 3.3, ensemble or hybrid approaches will be discussed.

3.1. Machine Learning-Based Techniques

Annuzzi *et al*. [9] used a Feedforward Neural Network (FNN) for predicting postprandial BGLs by using detailed nutritional factors such as carbohydrate and protein intake, among others. The model showed the importance of nutritional data, specific to patients, for precise predictions. However, its dependence on predefined nutritional combinations limits its adaptability to real scenarios where nutritional data might be incomplete or reported inaccurately.

Similarly, Annuzzi *et al*. [8] in another study expanded on [9] by considering ML methods to predict BGLs at a number of time horizons (15, 30, 45, and 60 minutes). Even though the study reinforced the role of detailed nutritional factors, it used a similar FNN approach, which, while effective, did not use advanced techniques like online learning or adaptive algorithms; this limited its applicability in real-time and dynamic environments.

In addition, another study by Bunescu *et al*. [16] adopted a unique approach by using physiological modeling with Support Vector Regression (SVR). This method offered informative features achieved better results compared to domain experts. However, its dependency on physiological models might not really consider for complexities such as sudden changes in lifestyle or psychological factors, which can affect glucose dynamics. Meanwhile, Perez-Gandia *et al*. [65] focused on CGM data with the use of an Artificial Neural Network model (ANN) for glucose prediction in real-time at different horizons (15, 30, and 45 minutes). This approach showed better accuracy than an autoregressive model but faced challenges with delays in prediction, especially for different downward trends. The dependence of this study on limited datasets from specific CGM systems may limit its applicability.

Anjum *et al*. [7] introduced an AI-assisted CGM system that used XGBoost, Seasonal Auto-Regressive Integrated Moving Average (SARIMA), and Prophet models for time-series analysis. This approach personalized interventions by identifying glucose trends and changing diets as needed. The use of multiple models improved predictions but its dependence on commercially available AI for the recognition of food raises concerns about both accessibility and consistency. The study also lacked validation on diverse cohorts of patients, which decreased its generalizability. In contrast, Elhadd *et al*. [27] evaluated glucose prediction in fasting patients in Ramadan with the use of XGBoost and achieved an R^2 of 0.836 and an MAE of 17.47. This model predicted normal and hyperglycemic events effectively but underperformed for hypoglycemia (27.9% accuracy). Combining glucose data and demographic and Physical Activity (PA) shows the importance of using diverse features, but the small sample size (n=13) and focus on a specific cultural practice decreases its applicability to larger populations.

A generic physiological model with SVR anticipated 25% of hypoglycemic events 30 minutes in advance but with only 42% precision; while it outperformed human experts, the high rate of false alarms and its focus on hypoglycemia highlight the need for improvement in feature selection and integration [66]. Moreover, its applicability in actual settings with glucose fluctuations is uncertain.

When comparing techniques, two studies Anjum *et al*. [7] and Elhadd *et al*. [27] highlighted the strength of time-series analysis and models like XGBoost. which had a number of effective features in addressing BGL variability. However, Elhadd *et al*. [27] used PA and demographic data while Anjum *et al*. [7] focused only on dietary adjustments, which shows a gap in detailed intervention strategies. Similarly, another two studies Medanki *et al*. [58] and Plis *et al*. [66] showed physiological modeling and closed-loop control systems as effective approaches for prediction in real time. Plis *et al*. [66] used the generated features from the physiological model of BG dynamics to train the SVR. However, Plis *et al*. [66] placed emphasis on prediction over action while the Medanki *et al*. [58] implemented effective interventions.

A different study in Rashid *et al*. [72] explored AI in the management of chronic diseases with the use of electronic health records and wearable devices. Their unique algorithms were designed for fairness in managing imperfect data. However, the paper focused more on the preprocessing of data and offered a limited quantitative comparison of prediction accuracy across models. Meanwhile, Basile and Sannino [12] highlighted the importance of interpretable AI in T1D by showing that Decision Trees (DTs) achieved the best F1 score (0.87) for predicting BGLs with the use of variability features of heart rate. The potential of non-invasive glucose monitoring is introduced in this work, but its classification approach oversimplifies the continuous data about glucose, which can affect prediction in the real world.

An ANN using time-domain features achieved higher RMSE values for different prediction horizons Alfian *et*

al. [2]. This study showed the importance of feature engineering for better accuracy, but it tested the model on a small dataset of only 12 patients, which decreases its applicability to large populations.

McShinsky *et al*. [57] investigated the performance of several machine learning models for glucose prediction and for detecting hypo- and hyperglycemic events over 30- and 60-minute prediction horizons. Their approach incorporated CGM data, wearable sensor readings, and self-reported events via mobile applications, when available. Among the models evaluated, the Least Absolute Shrinkage and Selection Operator (LASSO) regression achieved the best results, with average RMSE values of 20.58 mg/dL (1.14 mmol/L) and 33.41 mg/dL (1.86 mmol/L) for the 30- and 60-minute horizons, respectively.

Zhang *et al*. [95] evaluated four data-driven models for glucose prediction over 30- and 60-minute horizons. Each model was trained with varying feature sets depending on availability per subject. The Seq-to-Seq Long Short-Term Memory (LSTM) model achieved the best performance at the 30-minute horizon with an RMSE of 15.46 mg/dL (0.86 mmol/L), while Multiple Linear Regression (MLR) performed better at the 60-minute horizon, achieving an RMSE of 21.59 mg/dL (1.2 mmol/L) with lower computational cost. While MLR offers model interpretability, Seq-to-Seq LSTM lacks this advantage.

In Georga *et al*. [34], a patient-specific Random Forest (RF) model was trained on multi-input data (CGM, insulin, meal absorption, activity, etc.) for short-term glucose forecasting. The RF achieved average RMSEs of 6.60, 8.15, 9.25 and 10.83 mg/dL at 15, 30, 60 and 120-min horizons. A strength of this approach is the ability of RFs to handle nonlinear interactions among multiple inputs without assuming a model form. However, the model still had rising error with longer horizons, and its feature inputs (e.g. plasma insulin, meal absorption model, energy expenditure) may not be readily available in real-time. In a follow-up, Georga *et al*. [33] used a multivariable SVR on similar inputs. The SVR gave lower prediction error: average RMSEs of 5.21, 6.03, 7.14 and 7.62 mg/dL at 15, 30, 60, 120 min, respectively. This improved accuracy over [34] demonstrates that SVR can better exploit the same inputs. The strength of [33] lies in SVR's capacity to model complex relationships with kernel functions; its weakness is similar to [34]in requiring rich multivariate data (including accurate meal absorption) and being "offline" (training per patient). In summary, [34] and [33] show that adding more physiological inputs and choosing an appropriate algorithm (SVR over RF) can reduce RMSE by ~20-30%, but both require fine-grained input data and may still struggle with real-life variability.

Pappada *et al*. [63] used a standard ANN with multi-input (CGM, insulin, meals, activity, stress) to predict 75 min ahead. Its performance was relatively poor: RMSE was 43.9 mg/dL on the test set. It detected 88.6% of normal-range readings and 72.6% of hyperglycemia, but only 2.1% of hypoglycemia. In the Clarke error-grid, 92.3% of predictions fell in A/B zones (clinically acceptable). The key strength of [63] is demonstrating a real-time proof-of-concept with many inputs, but its weaknesses are stark: high RMSE and severe under-prediction of hypoglycemic lows, likely due to insufficient training examples of hypoglycemia. It contrasts with later works: e.g. Dave *et al*. [23] achieved >91% sensitivity and specificity for 30 min hypo prediction, showing that with better modeling (and perhaps more focused feature use) one can do much better than [63]'s naive ANN. In other words, [94]'s early ANN had inadequate accuracy for reliable use, illustrating that simple ANNs without task oriented tuning yield weak performance.

Zecchin *et al*. [93] described a "jump neural network" for 30 min forecasting, trained on two months of CGM and meal data. They report that this network was "feasible and sufficiently accurate", but do not give numeric errors. The model's strength is integration of past CGM and explicit meal timings, and it was tuned on real patient data. Its weakness is the lack of reported metrics makes it hard to assess; qualitatively it seems to have produced reasonable short-horizon forecasts but likely similar limitations to [63] (sensitivity to meal-reporting quality and limited hypoglycemia exposure). In general, [93] represents a step toward more sophisticated ANNs, but not clearly superior to earlier methods in accuracy.

Zecchin *et al*. [94] investigated adding a meal-absorption model to an ANN forecaster. They combined a ANN with a polynomial and physiological sub-model of meal carbohydrate appearance. The authors found that including meal information improved short-term prediction accuracy. While exact error numbers are not given in the abstract, this study's strength is demonstrating quantitatively that feeding carbohydrate data into the predictor reduces forecasting error. Its weakness is practical: it requires accurate real-time meal entry and a good sub-model of meal kinetics. This approach aligns with [33]'s multi-input SVR finding that more physiological inputs help accuracy. Overall, [94] shows that modeling the meal-to-glucose process (not just CGM alone) is beneficial, but it does not by itself achieve revolutionary accuracy.

Kushner *et al*. [47] took a very different tack: they built patient-specific "shallow" neural networks using a physiologically-informed structure, and made predictions up to 4 hours. Their networks achieved 60-min RMSE ≈28 mg/dL (plus 90-min: 33±4, 120-min: 38±6, 180-min: 40±8, 240-min: 43±12 mg/dL). Over these long horizons, >93% of predictions were Clarke-acceptable. The strength of [47] is the ability to forecast many hours ahead with reasonable accuracy by tailoring the model to each patient and constraining it with physiology. Its errors (e.g. 28 mg/dL at 60 min) are actually lower than those of many data-driven

population models at similar horizons. A weakness is that it requires building a new ANN per patient (needing individualized data), and performance degrades as horizon lengthens. Compared to [47, 48, 60] shows that careful personalization can match or beat generic DL: e.g., [47]'s 60-min RMSE (28) is better than ARISES's 35 or [60]'s 33 for open-loop; and its long-horizon forecast is unique. However, [47] is limited to patients with rich self-tracking (MDI+sensor).

Jensen *et al*. [45] explored predicting nocturnal hypoglycemia from CGM. Using CMG data and a forward-selection method, they built a logistic model. The result was an Area Under the Receiver Operating Characteristic Curve (ROC-AUC)of 0.79, with 75% sensitivity and 70% specificity. This is a modest performance, indicating hypoglycemia can be somewhat anticipated from CGM patterns. The strength is a simple model with a handful of features (forward selection yielded 4 predictors). The weakness is the limited accuracy (0.79 AUC) and small sample; compared to [81] or [23], this is relatively weak. For example, [81] attained AUC 0.966 for postprandial hypo with RF, showing that for some hypo tasks far higher accuracy is possible. In sum, [45] is a proof-of-concept with mediocre metrics, highlighting the difficulty of predicting hypos solely from CGM traces.

Favero *et al*. [24] did not introduce a predictive model but proposed a new metric Glucose Mean Squared Error (gMSE)to evaluate glucose predictors, which weights errors by clinical risk (overshoots vs undershoots). They showed that using gMSE instead of plain MSE can shift model selection toward predictors that better avoid dangerous errors. The strength of [24] is framing accuracy in clinical context. As a method, it complements the others (and Faccioli *et al*. [31] below explicitly uses it). It has no "accuracy stats," but it underscores that two models with identical RMSE may differ in safety. A weakness is that it's an evaluation tool rather than a predictor. In comparisons, [31] will later show that combining gMSE with specialized alarms can significantly improve hypoglycemia forecasts.

Faccioli *et al*. [31] combined Favero's gMSE with a "prediction-funnel" alarm strategy to forecast hypoglycemia online. Using a CGM test set, they report the best scheme had sensitivity (recall) 88% and precision 65% (F1 ≈75%) for hypoglycemia alarms, with only 0.29 false alarms per day and average 15 minutes of warning. In other words, most true hypos were caught (88%) while keeping false positives low. The strength of [31] is its focus on clinical usefulness: by weighting errors (gMSE) and requiring consistency (prediction funnel), it reduces missed hypos. Its weakness is remaining moderate precision (65%) and being tested offline. Comparatively, [23] achieved even higher sensitivity (~91%) and specificity (>90%) but in a different framework, and [31] has fewer false positives per day. Thus [31] shows how smart cost functions and alarms can improve pure predictive models for hypoglycemia detection, though there's still room to refine thresholds.

Cameron *et al*. [17] developed an "Embedded" Model Predictive Controller (EMPC) for an AP that explicitly minimizes glycemic risk. In silico, EMPC yielded a much lower risk index (BGRI ≈2.51) than standard MPC (≈3.05) or Pima Indian Diabetes (PID) (2.99) controllers, achieving 84% of glucose values in the safe range. The strength of [17] is its theoretical rigor: by embedding risk minimization and uncertainty into control, it better kept BG in range (84% in 70-180 mg/dL) than before. Its weakness is that it's a control algorithm, not a pure prediction model, and its results are simulation only. However, [17] demonstrates that even "AI" controllers can be tuned for patient safety (risk), reinforcing Favero [24]'s idea that how we measure success is crucial. We note [24] shows the value of risk-aware design, but it does not directly report forecast RMSE or similar stats to compare with the others.

Prendin *et al*. [68] performed a large head-to-head benchmark: they compared linear models (autoregressive; autoregressive moving-average; and autoregressive integrated moving-average known as Autoregressive Integrated Moving Average (ARIMA) versus nonlinear models (SVR, ANNs, etc.) for 30 min CGM forecasting (using CGM only). They found that an individualized ARIMA model achieved "best overall" performance with RMSE ≈22.15 mg/dL and hypoglycemia recall 82%, precision 64%. Their best nonlinear population model, ANN, had RMSE 21.81 mg/dL but only ~27% hypoglycemia recall. An individualized SVR achieved ~59-63% hypoglycemia recall with ~70% precision. In short, [68] showed that simple linear forecasting can rival (or even exceed) many black-box models, especially once hyperparameters are tuned per patient. A key finding is that the best model for glycemic safety may be an ARIMA, not necessarily a deep net. Their strength is the comprehensive comparison on real data; their weakness is limiting to CGM-only features (so meal/insulin info could change results). Compared to [48, 60, 68] suggests that the advantage of deep nets over classic time-series models is modest when individually tuned. This contrasts with work in [60] conclusion (LSTM best), but differences likely stem from [60] strong baseline tuning and metric focus on hypoglycemia detection rather than pure RMSE.

D'Antoni *et al*. [20] proposed an autoregressive time delayed jump neural network (ARTiDe-JumpNN) for BG forecasting using only past glucose. They report it "outperforms" other CGM-only models (in RMSE and hypoglycemia detection) on a 33-patient dataset, though no specific numbers are given in the abstract. The strength claimed is high accuracy with minimal inputs (CGM alone) and very little training data. Without reported RMSE we cannot compare quantitatively, but qualitatively this method seems promising as a lightweight alternative. Its weakness is that evaluation

details are opaque; it was benchmarked against general time-series and some literature models, but not against the latest deep nets. In context, [20] suggests that clever network design can squeeze more from CGM alone, but since it's CGM-only it likely underperforms multi-input approaches like [33] or [94].

Amar *et al*. [6] introduced multiple BG prediction models and indicated outstanding clinical accuracy 99.3% for 30-minute and 95.8% for 60-minute predictions with a reduction of 60.6% in severe errors, which was a clear sign that most of 30-minute predictions were clinically safe. Studies carried out by them reflected a variety of methods, which included a standard autoregressive model, a baseline model, and also more progressive ones such as tree-based methods (RF and light gradient boosting machine (LightGBM)), standard fully connected neural networks, and a novel Gradually Connected Neural Network (GCN) architecture that was built to give priority to recent data. The tree-based methods, while obtaining better numerical accuracy, did not perform very well in terms of clinically evaluated performance. On the other hand, all GCN models were superior in terms of clinical correctness, with GCN3, a variant of the GCN, being the best option as it significantly lessened the perilous clarke error grid predictions and demonstrated robust performance for patients with high glucose variability or those on the threshold of having hypoglycemia. The models made from silico data were even superior to those trained on real-world data. To sum up, GCN especially GCN3 possesses the best clinical utility for safe BG prediction in T1DM and beats other studies as well (e.g., 99.3% vs. 99.6% at 30 minutes in [60]).

Seo *et al*. [81] applied ML to predict postprandial hypoglycemia (triggered by meals). Using RFs on CGM and insulin data, they achieved AUC 0.966, with 89.6% sensitivity and 91.3% specificity for detecting an impending <70 mg/dL event. This is a very high accuracy. Strength: the RF found patterns in post-meal dynamics that enable nearly perfect discrimination (AUC0.97). Weakness: F1 score was only 0.543 (reflecting class imbalance). Relative to [45] (AUC0.79) or [74] (sens≈78%), [81] is far superior likely because it focuses on daytime postprandial hypos with supervised labels, where data are richer. It demonstrates that ML can effectively predict a specific event type (meal hypo) when appropriate features are used. Compared with [23, 31] (nighttime hypo), [81]'s better performance owes to clearer meal onset cues in CGM/insulin logs.

Parcerisas *et al*. [64] tackled the issue of nocturnal hypoglycemia in T1D patients undergoing Multiple Daily Injections (MDI) by utilizing CGM as well as other physiological variables. A variety of ML algorithms were tested, including linear discriminant analysis (LDA), LSTM, support vector machines (SVM), neural networks, naïve Bayes, and AdaBoost, with SVM obtaining the highest prediction accuracy and being chosen for further development of the model. Both the general and the personalized SVM models were developed with and without PA features through nested cross-validation, resulting in median sensitivities from 70 up to 75% and specificities from 72 up to 77%, with PA enhancing performance. These prediction models were then incorporated into implemented strategies that were evaluated using a modified UVA/Padova simulator, revealing the fact that the intake of 30g of rescue carbohydrates during hypoglycemia prediction leads to a reduction of nocturnal hypoglycemic occurrences by around 35-40%, a decrease of time below 70 mg/dL by ~40%, an increase of time-in-range by roughly 1.3%, all with results indicating statistically significant improvements. In conclusion, the research implies a substantial drop in nocturnal hypoglycemia along with the use of T1D patient safety sleep management through SVM prediction, targeted carbohydrate intake, and the combination of those practices.

Stuart *et al*. [84] used Logistic Regression (LR) on hospital EHR to predict inpatient hypo (<72 mg/dL). On 9584 admissions, they achieved AUC 0.733 (95% CI [0.719-0.747]) for predicting any hypo. This is modest discriminative power. Its strength is the very large clinical dataset and use of routine admission data (age, meds, labs). Its weakness is limited accuracy (AUC≈0.73) suggesting that static factors alone are insufficient. In contrast, Ruan *et al*. [78] (below) achieved AUC 0.96 using more sophisticated ML on in-hospital data, showing how model choice matters. Thus, [84] illustrates a baseline logistic model; it underlines that more advanced ML can substantially improve inpatient risk prediction.

Bertachi *et al*. [15] used CGM plus wearable activity monitors (accelerometers) to predict nocturnal hypoglycemia in T1D on MDI. Their SVM classifier achieved sensitivity 78.75% and specificity 82.15% for nocturnal hypos. Over 70% of nocturnal hypos could be avoided with their method. The strength of [15] is its multi-modal approach (CGM+movement), which outperformed CGM alone (as in [45]). The weakness is still~20% misses and false alarms, and data were collected in a specialized setting. [15]'s performance roughly matches [31]'s 88% recall (sensitivity) and [64]'s 79% recall. This suggests that adding activity data gives a slight edge over CGM-only models like Jensen [45] or Stuart [84], but is not a panacea.

Mosquera-Lopez et al. [61] took a big-data plus SVR model for nocturnal hypoglycemia. In silico simulations, their model, which used a decision-theoretic approach predicted 94.1% of nocturnal hypos (<70 mg/dL) with ROC AUC 0.86. They also analyzed insulin dosing to make recommendations. The strength is very high sensitivity (94%) to hypos, higher than most studies, at the cost of AUC=0.86 (moderate overall accuracy). Being in silico, its weakness is applicability to real patients. Compared to [15] (79% sens) and [23] (Dave's 91% sens), [61] shows that aggressive prediction rules

can catch most events (94%), albeit with an AUC similar to [45]. It reinforces that some approaches prioritize catching hypos (high recall), accepting more false alarms.

Ruan *et al*. [78] applied XGBoost on large EHR records to predict risk of clinically significant inpatient hypoglycemia (≤2.9 mmol/L). On~17,658 patients (32,758 admissions), the best ML model attained AUROC 0.96, far surpassing LR (0.75). This demonstrates ML's power on rich EHR data. Strengths: huge dataset, many features (meds, labs, vitals, history) and excellent discrimination (96% AUC). Weakness: retrospective design and single healthcare system, so real-world utility needs testing. [78]'s AUC 0.96 is the highest among these hypoglycemia-prediction studies, showing that in the hospital setting (with extensive data) ML can be extremely predictive. Compared to [84]'s 0.73 logistic, [78]'s XGBoost shows how nonlinearity and ensemble methods can unlock accuracy in big-data settings.

Guemes *et al*. [35] used multiple binary classifiers (SVM, RF, etc.) to predict "overnight control quality" (target vs out-of-range) using daytime CGM, meals and insulin in the OhioT1DM dataset. They report that no model was clearly best and that AUC≈0.70 was achieved. This modest AUC means only fair accuracy. Strength: innovative framing of the problem (binary good/bad night) using commonly collected features. Weakness: low AUC (~0.7), likely due to small dataset and the intrinsic difficulty of predicting with only daytime inputs. Compared to others, [35]'s results are relatively weak (e.g. [81] had 0.966, [31] had F1 75%). It suggests that predicting quality of control is much harder than predicting actual BG values with CGM.

Oviedo *et al*. [62] tackled postprandial hypoglycemia with supervised ML (features from CGM and context). They achieved median specificity≈79-81% and sensitivity≈71-77% for level-1 and level-2 hypos. So roughly 3/4 of hypos were caught, with ~80% of non-hypos correctly identified. This is comparable to [15] and [31]. The strength is use of real-meal data and stratified risk levels. The weakness is still significant misses (~20-30%) and it's an offline study. In context, Oviedo's sens/spec are slightly lower than [81]'s 90%+and [23]'s 91%. Overall, [62] confirms that event forecasting is feasible but far from perfect, needing trade-offs between sensitivity and false alarms.

Bertachi *et al*. [14] combined physiologic models with ANNs for BG and nocturnal hypo prediction. Although no stats are provided here, they presumably achieved moderate success. The concept's strength is using mechanistic glucose modeling to inform the ANN. Without data we can only note it aligns with [94] in blending physiology and ML, and likely had similar weaknesses (model misspecification, data needs). Since [14] is not fully published, we move on.

Reddy *et al*. [74] addressed a special case: predicting hypoglycemia at the start of exercise. They developed two algorithms: a simple DT "180/120 rule" and a more complex RF. The simple rule (if HR>121 bpm in first 5 min and start-glucose<182 mg/dL) achieved 79.6% accuracy, while the RF reached 86.7%. The strength is very clear, actionable rules for exercise, with higher accuracy for complex model. Weakness: these models were only for predicting during exercise initiation, a narrow task, and validated on limited data (12 patients). Compared to the rest, [74] is uniquely focused: for its domain it works well. It demonstrates that even simple ML (DT) can yield effective heuristics in specific scenarios (exercise).

Sudharsan *et al*. [86] tackled hypoglycemia prediction in T2Ds using Self-Monitored Blood Glucose (SMBG) records (very sparse) and basic ML. The study evaluated four ML models, including RF, SVM, K-nearest neighbor (KNN), and Naïve Bayes classifiers, where both RF and SVM achieved the strongest overall performance, exceeding 90% prediction accuracy across datasets. Their model (SMBG only) achieved 92% sensitivity and 70% specificity for any hypoglycemia in the next 24h. Incorporating medication data raised specificity to 90%. The strength is high sensitivity despite only using a few daily glucose readings. The weakness is the very low data density and wide 24h window so this is a much easier problem (spot hypo risk) than minute-by-minute forecasting. Compared to T1D CGM-based models, [86] operates in a different context. In sum, it shows that even with minimal inputs one can flag hypoglycemia risk with good sensitivity, at the cost of many false alarms (specificity 70% without meds).

Alhmiedat and Alotaibi [3] studied various ML algorithms for T2D prediction. These algorithms included LR, KNN, DT, Gradient Boosting (GB), RF, SVM, Light GB, CatBoost, ANN and Convolutional Neural Network (CNN). They used two different diabetes datasets where one of them used for training these models (when applicable) and the other for testing. Both datasets we preprocessed to enhance their system's reliability. They showed that the best resulted were obtained from RF with a classification accuracy of 98.95%. In contrast these findings, the authors in [5] who used the same dataset and almost the same ML techniques, found that LR and ANN were the best algorithms in predicting T2D in most evaluation criteria.

3.2. Deep Learning-Based Techniques

Eskaf [29] introduced advanced methodologies using genetic algorithms, DL, and Reinforcement Learning (RL) for predicting BGLs in the short-term. These techniques used real-time data for improving the accuracy of predictions and proved more adaptable. However, this study's dependence on simulated patient data and a small cohort of real patients affects their generalizability and also shows the need for validation on diverse people.

Prendin *et al*. in [69] placed emphasis on the

importance of model interpretability in T1D management through the application of SHAP, a tool for explaining black-box models, analysis to LSTM models. Though p-LSTM proved better in learning physiological relationships and improving glycemic control, the dependence of the study on retrospective data may not fully show conditions of the real world, where unexpected variability is capable of influencing accuracy.

Wang in [90] carried out a comparative study of ML techniques and found that bi-directional Bidirectional Long Short-Term Memory (Bi-LSTM) had the best performance for predicting hypo- and hyperglycemia (RMSE=7.55 mg/dL). Though the results are promising, the exclusive use of in-silico data decreases the generalizability to real-world scenarios in which there is more variability.

Askari *et al*. [11] studied event detection in diabetes management using deep learning techniques. They explored different Recurrent Neural Network (RNN) models and found that LSTM with 1D convolution layers and Bi-LSTM with 1D convolution layers provided the best performance for detecting meal and exercise events affecting blood glucose concentration (accuracy=94.72%, F1 score=96.85%). While these models showed high accuracy, the study relied solely on real-world data, which introduced challenges such as missing values and incomplete self-reported events. This limitation may impact the generalizability of the results, particularly in more diverse real-world settings where additional wearable sensor data could improve detection accuracy.

Zhu *et al*. [96] proposed a Dilated RNN (DRNN) for short-term 30-min glucose forecasting, achieving improved accuracy over traditional RNN models. Their method involved a transfer learning approach to enhance model generalization and performance. The DRNN demonstrated robustness and precision, with an average RMSE of 18.9 mg/dL (1.05 mmol/L). However, while the model achieved promising results, the study lacked clarity regarding the total number of instances used, which may affect the reproducibility and scalability of the approach in broader settings.

Rodriguez Leon *et al*. [76] evaluated the performance of three LSTM-based neural network architectures for forecasting BGLs in individuals with T1D. The models were trained for various prediction horizons 30, 60, 90, and 180 minutes and their performance was assessed using Clarke Error Grid analysis. All models achieved over 95% of predictions within zones A and B, indicating high clinical accuracy. The models were trained and tested using real-world samples collected from a small group of individuals.

Zhu *et al*. [97] developed the Adaptive, Real-time, and Intelligent System to Enhance Self-care (ARISES) system: a deep neural network using CGM, user-logged meals and insulin, and wearable sensor data (heart rate, EDA, etc.). They report (for 60-min horizons) RMSE ≈35.28±5.77 mg/dL, with Matthews correlation coefficients of 0.56 for hypoglycemia detection and 0.70 for hyperglycemia. Including the wristband signals improved RMSE by ~2.3 mg/dL. Compared to earlier deep models [48, 60], ARISES's RMSE is higher (worse)-for example, [60] had ≈20 mg/dL RMSE at 30 min (though at a shorter horizon). However, [97] uniquely shows that adding physiological signals (stress/activity via wearables) can modestly enhance predictions. Its strengths are the rich multi-sensors input and real-time deployment on a smartphone. Its weaknesses include a small sample size (12 subjects) and only moderate forecasting accuracy (RMSE ~35 at 60 min is quite large). It does outperform the old [63] ANN (43.9 mg/dL), but not the newer LSTM/RNN methods. In summary, [97] suggests that wearable signals yield incremental gain, but core forecasting remains challenging.

Li *et al*. [49] presented "GluNet," a deep CNN framework. They report that GluNet achieves state-of-the-art RMSE for 30- and 60-min forecasts, outperforming previous ANN/ARX/SVR models. On virtual adults and two clinical sets, they claim their RMSEs are the best reported. Though the exact numbers aren't listed here, this suggests 30-min RMSE in the rang~15-20 mg/dL and 60-min~30 mg/dL (comparable to [48, 60]). The strength is high accuracy via deep feature learning; the paper highlights significantly lower RMSE than e.g. NNPG, SVR, ARX. A weakness is again complexity: GluNet has many convolutional layers and needs substantial data. It's similar to [48] in theme: the best predictive accuracy so far on benchmarks. Compared to [49, 60]'s CNN may slightly beat LSTM alone; compared to [68], both find that tuned deep nets can edge out linear models in pure RMSE, albeit with heavy resource use.

Mohebbi *et al*. [59] compared recurrent networks (LSTM, GRU) to ARIMA for short-term CGM prediction. Their population LSTM (with pretraining) achieved 30-min RMSE ≈1.16 mmol/L (≈21 mg/dL) and 60-min RMSE ≈2.03 mmol/L (≈36.6 mg/dL). This is in line with [48, 60] (≈20 mg/dL at 30 min, ~33 mg/dL at 60). By contrast, a simple ARIMA or a patient-specific RNN had higher error at longer horizons. They concluded that population-trained RNNs outperformed patient-specific ARIMA ($p \leq 0.05$) for 60-min forecasts. Strength: carefully shows that shared "meta" models can generalize to new patients. Weakness: dataset small and limited to 15 patients. Compared to [47] (patient-specific shallow net), [59] shows that a generic LSTM can match or surpass ARIMA when multiple patients' data are pooled, reinforcing [60]'s finding that deep RNNs have an edge with enough data.

Daniels *et al*. [21] explored multi-task (meta-learning) neural nets for personalization. Using a Meta-LSTM approach with a few patient-specific examples, they achieved 30-min RMSE 18.8±2.3 mg/dL, 45-min 25.3, 60-min 31.8, 90-min 41.2, 120-min 47.2 mg/dL,

with ≥93% of predictions Clarke-acceptable at each horizon. This is among the lowest RMSEs reported (e.g. 18.8 mg/dL at 30-min beats most others). The strength is very high accuracy combined with personalization even with limited data (meta-learning). Its weakness is the complexity of MAML training and the need to select tasks. In comparison, [47] (patient ANN) had 28 mg/dL at 60 min, [60] had 19.8 at 30 min. [78]'s 30-min (18.8) is better than [48, 60]; its 60-min (31.8) roughly matches [48, 60]. Thus, [21] shows that advanced meta-learning can push accuracy to new highs, rivaling or exceeding state-of-the-art.

3.3. Ensemble or Hybrid Techniques

Sampa *et al*. [79] employed a Gradient Boosting Decision Tree (GBDT) for predicting BGLs among Bangladeshi workers with the use of non-invasive health checkups and dietary data. With an RMSE of 2.30, this model achieved high accuracy and emphasized cost-effective monitoring of health for developing regions. However, the lack of CGM data and its dependence on static measures limit its responsiveness to glucose variability in real time. On the other hand, other research [58] proposed an insulin delivery system controlled by PID with AI for the regulation of glucose. This closed-loop system managed glucose spikes after meals in an effective manner and used personalized, real-time data for insulin administration. However, the focus on technological complexity such as WiFi-controlled micropumps poses challenges in settings with limited resources. In addition, this study did not evaluate the long-term efficacy of the system or the adherence of patients.

Srinivasu *et al*. [83] introduced an Explainable Artificial Intelligence (XAI) system using Bi-LSTM and CNN for classifying levels of glucose through spectrogram images. While the model showed efficacy in real-time monitoring and interpretability, it relied on strips of glucose oxidase, which are semi-invasive. This dependence decreases its potential.

Alkanhel *et al*. [4] developed a hybrid Gated Recurrent Units and CNN (CNN-GRU) model for glucose forecasting in real-time in Internet of Things (IoT)-based systems, which outperformed other models. IoT offers a number of advantages, but this study's dependence on public datasets highlights concerns about variability. Hybrid CNN-GRU [4] and ANN [12] achieved high accuracy but their datasets were different in size and nature. Bi-LSTM [72] excelled in RMSE but was limited by in-silico data; many studies relied on limited datasets and lacked validation [90] in actual settings. In addition, non-invasive monitoring methods were explored in only one study [12].

Martínez-Delgado *et al*. [55] proposed an RNN-LSTM model combined with various insulin and carbohydrate absorption curve configurations to predict BGLs in individuals with T1D. The input data included glucose, acceleration, insulin (basal and bolus), and carbohydrate intake. The authors conducted seven experiments by varying the absorption functions, with the best configuration achieving an RMSE of 9.2 mg/dL (0.510 mmol/L) for a 60-minute prediction horizon.

Mosquera-Lopez and Jacobs [60] tackled forecasting by training an LSTM- RNN on a very large real-world dataset. Using 175 "closed-loop" and 75 "standard" patients (thousands of days of data), the LSTM achieved RMSE ≈19.8±3.2 mg/dL at 30 min and 33.2±5.4 mg/dL at 60 min for the closed-loop data (similar RMSE ≈19.6 and 33.1 mg/dL for the open-loop data). Nearly 99.6% of 30 min forecasts fell in zones A+B of the Parkes error grid, indicating high "clinical accuracy." The main strength of [60] is the use of deep LSTM to capture long temporal patterns and the introduction of new metrics (the glucose variability impact index and prediction-consistency index) for evaluation. The LSTM outperformed simple ridge and RF baselines on this dataset. Its weakness is that even with a huge dataset, 30 min RMSE around 20 mg/dL remains relatively high for strict control, and LSTMs require substantial data and computation. Nevertheless, [60] demonstrated that deep learning on large CGM databases can improve forecasting relative to earlier models. Compared to [48] (below), [60]'s LSTM had slightly better RMSE (≈20 vs ≈21 mg/dL at 30 min), although [48] included different input handling.

Li *et al*. [48] proposed a hybrid convolutional-recurrent neural network (a "CRNN+LSTM") called GluNet. On the UVA Simulator (in silico), GluNet achieved 30-min RMSE ≈9.4 mg/dL and 60-min RMSE ≈18.9 mg/dL. On 6 real patients, it gave 30-min RMSE ≈21.1 mg/dL and 60-min ≈33.3 mg/d. These are comparable to [60]'s results on clinical data (≈20 and 33 mg/dL). The CRNN+LSTM automatically extracted features from the CGM time series and outperformed a variety of benchmarks (SVR, LVX, ARX, pure LSTM) on the simulated data. A major strength is its overall high accuracy on simulations and competitive clinical performance; additionally, it was implemented on a smartphone to show real-time feasibility. Its weaknesses include its complexity (many layers, need of GPUs), reliance on good training, and, on real data, only modest improvement over simpler methods. In [48], the CRNN+LSTM outperformed classic Autoregressive model (ARX) and Latent Variable model (LVX) at 30 min but ARX caught up at 60 min, indicating no clear "best" model for all horizons. Overall, [48] confirms that deep architectures can match or slightly beat classical algorithms, but it also illustrates that nonlinear CRNN+LSTM complexity may not dramatically lower RMSE on real CGM beyond ~20-30 mg/dL.

Dave *et al*. [23] built a real-time classifier for nocturnal hypoglycemia using engineered features (time-of-day, recent CGM trends, meal/insulin history). Their ensemble model achieved >91% sensitivity for both 30- and 60-min hypoglycemia prediction, with

>90% specificity. This is one of the strongest results: it predicts ≥91% of events, much higher sensitivity than [14, 15, 31, 62]. The strength is outstanding sensitivity and specificity in a large dataset; also feature-based models are interpretable. Its weakness might be reliance on multiple inputs (glucose, insulin, meals) and being offline. Compared to [31]'s 88% recall (and 65% precision), [23] yields both very high recall and precision. [23] thus represents a high-water mark for hybrid ML in real-time hypo alarms.

Sampath *et al*. [80] proposed aggregating multiple glycemic-control indices into a predictor for nocturnal hypos. On a CGM dataset, their ensemble achieved sensitivity 77% and specificity 83.4%, with PPV 80.2% and NPV 80.6%. This is comparable to [15, 62] (which had~78%/81%). The strength is combining multiple simple predictors to slightly improve performance. The weakness is that even the aggregated model is only~80% accurate–so many events are still missed. Compared to [23] (91% sens) or [61] (94% sens), these figures are lower, reflecting that [80] used only SMBG-based indices and no sophisticated learning. It shows that moderate accuracy can be achieved by aggregating prior simple methods, but falls short of advanced ML models.

In summary, these 57 studies illustrate a wide spectrum of AI techniques for glucose/hypoglycemia prediction. Early neural-net models [63] had high RMSE and poor hypo detection, but later SVR [33] and deep nets [21, 48, 49, 69] steadily reduced RMSE (to~20 mg/dL at 30 min). Models using additional inputs (insulin, meals, wearables) generally outperform CGM-only models: compare [33, 94] (multi-input SVR/ANN) vs CGM-only ARX/NARX. Hybrid approaches (e.g. [14, 94]) that combine physiology with ML often improve accuracy, but at the cost of added complexity. In hypoglycemia alarm prediction, recall can be pushed high (e.g.~90% in [23, 61]) by accepting more false alarms, whereas precision remains moderate unless models are very precise or narrow in scope. Linear time-series ARIMA remain competitive with complex nets when personalized, as [68] shows. Large-scale EHR models [78] and big-data LSTMs [60] achieve very high AUC or accuracy, but require massive datasets.

The strengths of deep/ensemble AI include flexibility with inputs and often improved average accuracy, while weaknesses include data hunger, complexity, and sometimes limited clinical interpretability. Simpler models (ARIMA, linear methods) can be surprisingly effective if well-tuned. For nocturnal or exercise-related hypos, specialized features (e.g. heart-rate rule [74] or activity sensor [15]) are crucial. Accuracy metrics span RMSE from ~9-43 mg/dL (depending on horizon and data) and AUC from ~0.70 up to 0.96. Notably, a "clinically accurate" classifier by Amar [6] claims 99% in-range forecasts, far above others-if validated, this suggests there is still room to learn from ensemble or hybrid approaches. In comparing results, it is clear that no single technique dominates all tasks: some methods excel at short-term CGM forecasting (deep networks), others at event prediction (RF for postprandial hypo), and some at broad risk stratification (XGBoost on EHR). The evolving landscape shows progressive gains in prediction accuracy over the years, but also persistent challenges: even the best models still miss hypoglycemic events or carry substantial RMSE. Each study's findings complement and sometimes contrast with others, painting a picture that careful choice and customization of AI methods-with attention to input data and clinical priorities-is essential for optimal glucose prediction.

Data obtained from these studies is summarized in the Table 1 and shown in Figure 5.

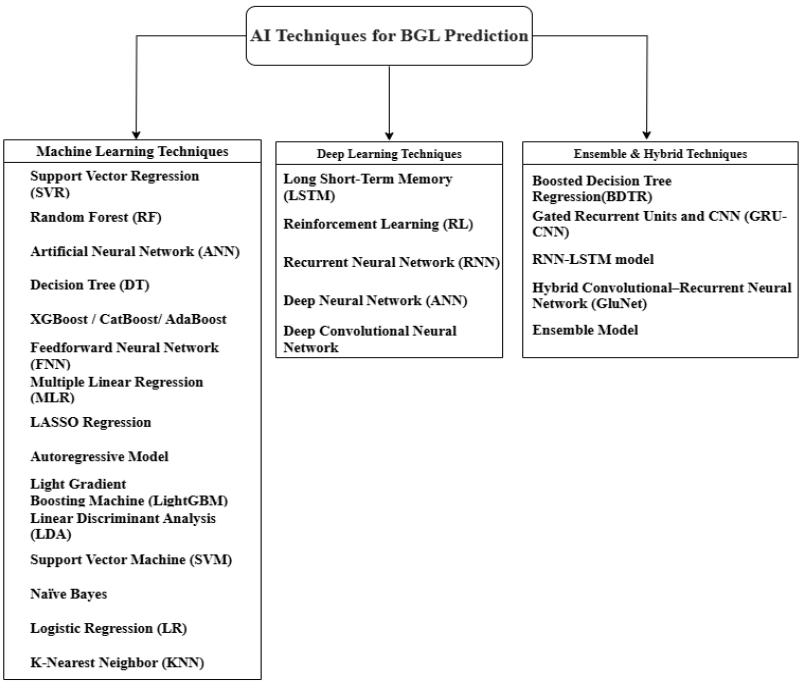

Figure 5. Taxonomy of AI techniques for blood glucose prediction.

4. Datasets

This section reviews the datasets used in the research papers analyzed in this study, providing access information for researchers interested in similar studies.

The authors of two studies [8, 9] used two different datasets: DirecNet [43] and AI4PG [46]. DirecNet is a public dataset of CGMs collected from pediatric patients with T1D via a wearable device that recorded glucose values at intervals of five minutes [43]. The dataset involved 50 participants divided into two age groups: ages 3 to less than 7 years and ages 12 to 17 years. The dataset encompasses detailed records of continuous glucose measurements, hormonal response data, and parental observations, providing valuable insights into the physiological and symptomatic aspects of hypoglycemia in pediatric T1D patients.

The AI4PG dataset contains data from 25 T1D patients [46]. This dataset spans six to seven days and includes information on meals, insulin doses, and CGM readings. The participants, aged 40±12 years, had an average diabetes duration of 15±12 years. Over a week, patients maintained detailed food diaries, resulting in 1,264 recorded meals (including breakfasts, lunches, and dinners), represented as time series of glycemic levels before and after meals. The dataset provides information on manual insulin boluses administered based on meal carbohydrate content, along with estimates of carbohydrate, lipid, protein, fiber, and caloric intake. CGM readings were recorded every five minutes, from 30 minutes before to 60 minutes after meals.

Other research [16] used a dataset involving both real and simulation-based dataset for predicting CGM in individuals with T1D. The real dataset was collected from five T1D patients, with CGM readings taken at five-minute intervals. This data includes information about daily events affecting glucose levels, such as insulin doses, meal intake, exercise, and sleep. The dataset consists of 200 timestamps manually selected from the dataset, with 40 points per patient. These timestamps reflect diverse situations, including different times of day or night, varying curve trends (rising, flat, or falling), and proximity to meals, insulin, or other events. The simulated dataset was used for hypoglycemia prediction; an additional dataset of 5,816 test points was constructed, containing 152 hypo events. Non-hypo points were randomly sampled to maintain the ratio observed in patient data. CGM readings below 70 mg/dL lasting over 20 minutes were considered hypo events.

Another dataset [65] included glucose monitoring data from two CGM systems: the Guardian® Real-Time CGM System (Medtronic-Minimed) and the FreeStyle Navigator® CGM System (Abbott Laboratories). The Guardian® system provided glucose measurements every five minutes (288 samples per day) and included data from nine T1D patients who wore the device intermittently for 72 hours per week over four weeks, contributing 12 daily profiles per patient (six full-day and six half-day recordings). The FreeStyle Navigator® system provided measurements every minute (1,440 samples per day) and includes data from six patients who wore the system for approximately 72 hours, contributing two complete daily profiles per patient. The dataset captures a variety of glucose profiles, including hypo- and hyperglycemic events, providing a comprehensive source for developing and testing glucose prediction algorithms.

One study Anjum *et al*. [7] focused on real-time CGM data collected from eight T2D patients over a period of six months. The CGM system provided glucose readings every 15 minutes for a duration of 15 days per device without requiring traditional finger pricks. The dataset included essential glycemic parameters such as mean glucose level, variability, and the duration of hyper- and hypoglycemic episodes. The study primarily used 10,160 glucose readings after preprocessing.

To account for fasting, the dataset used in the PROFAST-IT Ramadan study [27] comprised CGM and PA data collected from 13 patients with T2D during Ramadan. The participants, aged between 18 and 79, were on multiple glucose-lowering therapies, including sulfonylureas or insulin, and wore a Fitbit-2 pedometer along with the Freestyle Libre CGM sensor for two weeks before and during Ramadan. The dataset includes glucose measurements captured every 15 minutes and PA data aligned with CGM readings. The study used a total of 19,540 samples and 55 variables, including clinical, PA, and time-based features, for training and testing machine learning models.

Plis *et al*. [66] included CGM data collected from five T1D patients over multiple days. The dataset contains glucose readings at regular intervals (200 timespans and 40 points per patient) along with relevant contextual information such as insulin dosages, meal intake, PA, and other factors affecting BGLs. The dataset also includes a larger evaluation set of 5,816 test points for hypoglycemia prediction, constructed by combining consecutive CGM readings under 70 mg/dl into hypo events and sampling non-hypo points to reflect the true distribution of hypo and non-hypo readings.

Further work [58] primarily involved simulations using the Bergman Minimal Model to represent blood glucose dynamics. This model employs a set of differential equations to simulate the interactions between glucose and insulin in the human body.

Other data [72] involved both real-world and simulated data. For real-world data, the researchers utilized free-living data from individuals with T1D, including CGM, insulin pump usage, meal intake, and PA information. The dataset used is from the Tidepool database, which consists of data donated by 300 individuals with T1D [88]. Around 50 datasets within this database include detailed CGM, insulin pump, and PA information, with data lengths ranging from 30 to

700 days. Additionally, the study employs simulated data using the mGIPsim multivariable simulator, which generates glucose and insulin concentrations and simulates patient behavior for up to one year. This simulated data helps evaluate and test the effectiveness of advanced multivariable automated insulin delivery systems. The Tidepool dataset [88] was also used in [60].

One study [12] was derived from the D1NAMO dataset, a publicly accessible dataset available at [25]. It includes data collected over eight weeks from 29 subjects, comprising nine individuals with T1D and 20 healthy individuals. Data acquisition involved CGM and physiological signals such as ECG, breathing, and accelerometer recordings using the Zephyr BioHarness 3 band. The glucose levels for diabetic patients were measured every five minutes using the Medtronic iPro2 Professional CGM sensor, while healthy subjects used a Bayer Contour XT glucose meter. The study only focused on the data from three T1D patients, as the data for other diabetic participants was either misaligned or too noisy. The same dataset D1NAMO was also used in [76] with six T1D subjects divided into windows of 24 samples for a total of 2,740 windows.

Another study [2] focused on the DirecNet dataset [43], which includes real-world data from 12 children with T1D. The data was collected using CGM devices (Guardian-RT), which recorded BGLs every five minutes for approximately seven days. The dataset was processed using various filtering and normalization techniques to improve accuracy in predicting future BGLs.

Georga *et al*. [34] and Georga *et al*. [33] created a private dataset collected from twenty-seven T1D patients (12 females, 15 males), aged 19-72 years, all on multiple-dose insulin therapy. Data were collected over 5-22 days per patient under free-living conditions. It consisted of subcutaneous glucose measurements, Energy expenditure, meals data and insulin injections information.

Another study [93] supported their study by using a real-world dataset collected from 20 individuals with T1D. It was collected using DEXCOM SEVEN PLUS CGM sensors and it was not publicly available. The dataset included CGM and SMBG values, CGM trend information, insulin and carbohydrate records, hypoglycemic and hyperglycemic symptoms, physical activities, and even emotional factors. In [93, 94], the authors used CGM data collected from 20 T1D patients, along with simulated data generated for 9 additional patients using the UVa/Padova simulator. In [97], a real T1D dataset compromising12 patients was used as a test study.

Kushner *et al*. [47] supported their study using a real-world dataset collected from 24 T1D patients. The dataset included 16 patients using insulin pumps CSII while the remaining 8 patients used Multiple Daily Injections (MDI). Collected data included CGM and SMBG measurements, records of insulin delivery and carbohydrate intake, and physical activities.

A large real-world clinical dataset, collected from 463 T1D patients, served as the basis for the experiments used in [45] to validate their proposed technique for predicting nocturnal hypoglycemia patterns. The dataset covers approximately 4,700 nights of CGM data, enriched with meal, bolus insulin, and demographic information. The dataset is proprietary and not publicly available.

The study in [31] utilized a real-world dataset obtained from 11 T1D patients to validate their finding in predicting hypoglycemia patterns. The dataset contained various data types including CGM, insulin, and meal information, and can be found through [42].

A simulated dataset of 9 T1D patients, generated using the UVa/Padova simulator, was the core dataset used to validate the proposed algorithm in [17]. The dataset included glucose levels, insulin doses, and predefined meal inputs.

A real-world dataset of 124 T1D patients, collected using Dexcom G6 CGM sensors, was the core dataset used to validate the proposed algorithm in predicting hypoglycemia patterns [68]. As reported, the CGM data were the only data used in their experiments and dataset is not publicly available.

D'Antoni *et al*. in [20] used both private and public real-world datasets where the private dataset obtained from 33 T2D patients while the public dataset consisted of 6 T1D patients obtained from the OhioT1DM dataset [54].

Both simulated and clinical datasets for T1D were used by Amar *et al*. in [6] to evaluate their proposed algorithm. The simulated dataset consisted of 30 cases generated using the UVa/Padova simulator. The clinical dataset was a set of 141 real-world cases each containing glucose readings and insulin dosage data, while the simulated dataset had extra information regarding the meal (carbohydrate) data.

Seo *et al*. [81] used a real-world dataset collected from 104 T1D patients to evaluate the proposed algorithm in hypoglycemia detection. These patients were divided evenly into groups where the first group had T1D and the second had T2D. CGM data collected every 5 minutes using the Medtronic CGMS Gold™ device. The only other data used was the timing of meals, which were inferred from calibration finger-stick blood glucose measurements performed before eating. The study did not use data on insulin doses, carbohydrate intake, or exercise. Instead of using raw CGM data, the researchers engineered a set of features to train their models: Rate of Increase in Glucose (RIG): The speed of glucose rise from a meal to its peak. Glucose Rate of Change (GRC): The near-instantaneous change in glucose levels. Current CGM Value: The glucose level at the time of prediction.

A private clinical dataset collected from 10 T1D patients who are prone to hypoglycemia was used in the

experiments conducted in [15, 64] to evaluate their proposed ML models. Its main characteristic is the combination of CGM, PA tracking, and self-reported data, which was then heavily processed to create engineered features for the prediction models.

A large-scale dataset based on real-world, routine hospital data rather than specialized monitoring devices was used in [84] to predict hypoglycemia in hospitalized patients with diabetes. The dataset comprised 9,584 hospital admissions for 6,187 unique adult patients with a diagnosis of diabetes. Unlike other studies that rely on CGM, this work used routinely collected data available in standard electronic health records. The dataset included: demographics data such as Age, sex, ethnicity, and an index of deprivation. Also the admission information and the Medication History such as insulins or oral diabetes medications. Furthermore, it included laboratory results such as eGFR (kidney function), CRP (inflammation marker), sodium, albumin, hemoglobin, and more. Finally, the occurrence of hypoglycemia, defined as a blood glucose level of<4.0 mmol/L, measured either by point-of-care testing or a laboratory blood test during the hospital stay.

Mosquera-Lopez *et al*. in [61] used two real-world datasets: the first, from Tidepool [88], included 124 T1D patients and was used to train their proposed model, while the second, obtained from 10 T1D patients, was used to validate their proposed model. The datasets were collected from a variety of different CGM and insulin pump devices. The key information used was CGM readings (taken every 5 minutes) and insulin dosing information from the pumps.

A large-scale dataset of 17,658 in patients with diabetes, covering 32,758 hospital admissions obtained over four years from Oxford University Hospitals, was the building block of the validation process of the proposed model in [78]. The dataset contained a comprehensive demographic, clinical, laboratory, medication, and procedural information, used to predict hypoglycemia risk and unfortunately it is not publicly accessible.

The study in [62] relied on retrospective data collected from 10 T1D patients where it contained CGM data in addition to baseline data such as HbA1c levels, blood pressure, lipid profiles, and any other relevant medical history or demographic information.

Reddy *et al*. in [74] used a real-world dataset compiled from three different clinical studies and collected from 55 T1D patients using a close-loop system. The dataset contained anthropometric characteristics, exercise-related data (heart rate, energy expenditure), hormone therapy details (insulin on board, daily insulin dosage, glucagon use), and glucose levels at the start of exercise.

Sudharsan *et al*. in [86] used a real-world dataset collected from T2D patients by utilizing SMBG values. The dataset included multiple SMBG values per patient, with an optimal frequency of approximately 10 SMBG values per week. However, authors did not mention how many patients participated in the study.

Another study [29] employed two datasets to include both simulated and real-world data for predicting BGLs in T1D patients. The simulated data was generated using the SimGlucose system, an open-source tool that simulates glucose-insulin dynamics based on predefined patient profiles derived from clinical data. These profiles cover a range of age groups, including adolescents, adults, and children, with a total of 30 unique profiles. Real-world data was collected from three volunteers, including two diabetic individuals using the DexCom seven CGM system. This monitoring device provides glucose readings every five minutes and allows for real-life data collection under routine conditions without activity restrictions. Readings were recorded every five minutes over a continuous 24-hour period for seven days using both the SimGlucose and DexCom CGM systems. The dataset underwent preprocessing steps such as filtering, normalization, and feature extraction. Discrete Fourier Transform (DFT) was applied to determine the natural frequency and detect short-term blood glucose fluctuations. The dataset comprised 12 blood glucose level readings over a one-hour window, with the goal of predicting the blood glucose level 30 minutes later. This structure was then used to train and evaluate the predictive models.

Other research [69] was derived from the OhioT1DM dataset [54], which includes data from 12 individuals with T1D monitored over eight weeks. For this particular research, data from one specific subject (ID 588, a compliant female participant aged between 40 and 60 years) was selected. This participant consistently reported meal and insulin information, providing detailed data for analysis. The dataset contains CGM data, insulin infusion records, and other physiological parameters. These data points were collected using Medtronic Enlite CGM sensors and insulin infusion pumps. The same dataset OhioT1DM [54] was also used in [96], where glucose samples from six subjects were analyzed to evaluate their results. In addition to that, the study utilized a simulated dataset generated using the UVA/PADOVA T1D simulator, which included data from 10 virtual subjects comprising 103,680 samples. Furthermore, the same OhioT1DM dataset [54] was employed in [14, 21, 24, 35, 57 and 95] to train and test their proposed models for 30- and 60-minute prediction horizons using data from all 12 subjects.

Wang [90] used the PID dataset [82], which consists of 768 T2D female patients, each at least 21 years old, with nine unique data points: pregnancy count, Body Mass Index (BMI), insulin level, age, blood pressure, skin thickness, glucose levels, diabetes pedigree function, and an outcome variable indicating whether the person has diabetes. The same dataset [82] was also used by Alhmiedat and Alotaibi [3], with a focus on improvement data quality by addressing missing, incorrect, or inconsistent values. This is in addition to

their use of other dataset obtained from the hospital Frankfurt in Germany [22]. In the same context of dataset preprocessing, Aouragh *et al*. [10] focused on addressing data imbalance in the Pima dataset [82]. They evaluated three imbalance-handling techniques: Resampling (RESAMPLE), Synthetic Minority Over-sampling (SMOTE), and Adaptive Synthetic Sampling (ADASYN). Their finding showed that RESAMPLE performed best in mitigating the data imbalance problem. Ramalingaswamy and Kuppili [71] proposed an optimized RBF neural network for Pima dataset [82] classification , using cluster validity indexes to determine the optimal number of hidden neurons, reducing network complexity by 90%, and improving accuracy to 73.5%. It employs clustering, Gaussian basis functions, and genetic algorithms for weight optimization, addressing hidden layer size issues and enhancing classification efficiency.

One study [11] consisted of real-world data collected from people with T1D in a free-living environment. It includes CGM sensor data, insulin pump records, and self-reported meal and PA information. The study aims to detect unannounced disturbances in blood glucose concentration caused by meals and exercise. The dataset contains 300 self-collected T1D datasets, out of which 50 include CGM, insulin pump, and exercise information [32].

Another study [76] supports their finding by using a real-world dataset collected from individuals with T1D named as T1DiabetesGranada [77]. This dataset includes continuous glucose levels along with patient demographic and clinical information. It comprises 257,780 days of data collected over four years from 736 T1D patients in the province of Granada, Spain.

Three datasets were used in [49] to evaluate their proposed algorithm: two real-world datasets and one simulated dataset. The simulated dataset consisted of 10 cases generated using the UVa/Padova simulator. The first real-world dataset was a set of 6 real-world cases obtained from the OhioT1DM dataset [54]. The second real dataset was a set of 10 real-world cases obtained from an internal project. All datasets contained glucose readings, insulin dosage data, and meal information.

Fifty patients with diabetes were used as a real-world case study in the work done by [59]. The dataset contained, as in other studies in this domain, glucose readings, insulin dosage data, and meal information, and dataset is not publicly available.

Another study [79] utilized CGM data collected from 271 employees of the Grameen Bank complex in Dhaka, Bangladesh. It includes basic noninvasive health checkup measurements, dietary information, and sociodemographic characteristics. The health checkup data consists of clinical measurements such as blood pressure, pulse rate, body temperature, BMI, waist-to-hip ratio, blood oxygenation (SpO2), and BGLs. Dietary information includes the frequency of consuming sugar-containing drinks and fast food, while sociodemographic data covers age, sex, and education level.

Another dataset [83] focused on glucose levels collected using glucose oxidase sensors connected to the human body. The collected signal data is processed and converted into spectrogram images through Short-Time Fourier Transforms (STFT). These spectrogram images are labeled into classes of low, average, and abnormal glucose levels for training a CNN with Bi-LSTM model. The dataset comprises 6,735 spectrogram images, with 2,806 images for normal glucose levels, 2,658 for high glucose levels, and 1,271 for low glucose levels. The study emphasized classifying these spectrogram images to aid real-time monitoring of diabetic patients.

Further datasets [4] included a publicly accessible T1D dataset sourced from Kaggle [32]. It contains real-time BG measurements with different forecasting timeframes ranging from five to 60 minutes ahead from a single patient with T1D, structured as a time-series dataset.

Both simulated and clinical datasets for T1D were evaluated in the proposed algorithm by authors in [48]. The simulated dataset consisted of 10 cases generated using the UVa/Padova simulator. The clinical dataset was a set of 10 real cases each containing glucose readings (obtained every 5 minutes using Dexcom G4 Platinum CGM sensors), insulin bolus, and meal (carbohydrate) data. In the same context of using real clinical dataset, authors in [63] had an in-house dataset consisted of 10 T1D patients with glucose readings using CGM sensors along with other clinical information.

The study in [23] used CGM readings obtained from 112 T1D patients using Dexcom G6 CGM devices to validate their proposed ML models for real-time hypoglycemia prediction. In addition to CGM data, the dataset also included basic demographic and clinical characteristics of the patients, such as: gender, age, HbA1c, duration of Diabetes.

The study in [80] used two real-world T1D datasets named DIAdvisor dataset and ChildrenData dataset. The DIAdvisor dataset collected from 34 adult patients with four daily intermittent BG measurements using finger sticks and CGM data sampled every 5-10 minutes. The ChildrenData dataset was collected from 179 children with nine BG measurements taken at specific time points over a 24-hour period.

This review of research revealed a significant lack of publicly available real-world datasets that researchers can use. Most studies depended on unpublished or proprietary data, making it difficult for others to replicate or build upon existing work. While some published datasets do exist, they are often too small or limited in scope, particularly for machine learning applications, which require large and diverse datasets for effective training and evaluation [41]. This scarcity of accessible data poses a major challenge for researchers seeking to develop and test new algorithms. However, this does not prevent the existence of many attempts to form reliable

datasets that can support future research. Table 2 summarizes the reviewed datasets, and notes that the Jaeb Center for Health Research provides access to some datasets through its website [42]. Readers who are also interested in generating a simulated diabetes dataset can be referred to such as: UVA-Padova T1D simulator [53], OHSU T1D simulator [75] and others [30, 36, 73, 91].

5. Discussion

The systematic review presented here merges up the results of 57 studies published during the period from 2010 to 2025 and reveals a major change in the use of AI for BGL forecasting. The results emphasize the move from classical physiological modeling to cutting-edge ML and DL, which, though they have their different pros and cons, are still facing this challenge in clinical acceptance.

The review points out a very clear and distinct trade-off regarding model complexity versus interpretation. Among the latter, DL, especially LSTM networks and RNN, has proven best in the task of handling the time-dependent nature of data from CGM. Seq-to-Seq LSTM methods, for instance, have been able to achieve unbelievably low RMSE values of just 15.46 mg/dL in 30-minutes forecasting. Meanwhile, these "black box" models are unable to provide explanations that are essential for building trust in clinical settings.

The future is thus hybrid architectures, which have come up as the most attractive option. The CNN-GRU and CNN-LSTM pairs always get the highest predicting power because they are good at both interpreting the picture shown by the glucose-related features and catching the changing patterns over time in the glucose monitoring. Take, for instance, the CNN-GRU model proposed by Alkanhel *et al*. [4], which combines convolutional feature extraction with GRU to manage long-term dependencies and thus achieves good performance in both short-term and real-time forecasting. Likewise, models using both convolutional and recurrent techniques, such as CRNNs and CNN-LSTM frameworks, have shown better performance than the traditional RNNs by maintaining detailed temporal structure while at the same time alleviating the vanishing-gradient problem [48, 49, 96].

On the other hand, simpler ML models, such as DT, RF, and SVR, are more transparent and have been proven to be competitive in performance, especially using heart rate variability, F1 scores of up to 0.87, as an integrated feature. Their explainability and lower power-consuming nature make them ideal for wearable technology and point-of-care devices that operate on low power or are embedded systems. One of the most common things in the literature is that the use of CGMs gives a better prediction when combined with other auxiliary inputs. The studies that used meal composition (carbohydrate, protein), insulin, and PA together with trends of glucose continuously beat those that relied on them only. To illustrate, Zecchin *et al*. [94] showed that the forecasting error went down by about 18% due to the integration of the carbohydrate data. Also, event detection, like exercise-induced hypoglycemia, was improved by the models that included non-invasive signals such as heart rate or activity sensors.

This multimodal method has received support from interpretability techniques like SHAP, which have exposed the lifestyle factors that were not observed but still play a significant role in model uncertainty [69, 97]. The proof is very strong that the high-fidelity BGL prediction will be realized in multimodal sensor fusion and not in the isolated glucose tracking, as this addresses the missing context problem in real-world free-living datasets through the lens of paradigm shift in sensor technology and data processing techniques, respectively.

Hypoglycemic event prediction, notwithstanding the high overall accuracy of the models, continues to be a great challenge. For example, while XGBoost achieved over 95% accuracy for normal ranges, the performance on hypoglycemia detection dropped significantly, with some studies reporting the sensitivity as low as 27.9%. A common finding across different studies is the existence of a trade-off between very sensitive detection of hypoglycemia and low precision. Models that are fine-tuned to pick up the downward trends in glucose, which are crucial for timely hypoglycemia alerts, usually end up creating more false alarms, as minor fluctuations in CGM readings can pass off as early hypoglycemic signatures.

Prediction of hypoglycemia is very challenging because of the very fast physiological onset and the impact of external factors such as sudden physical exertion. This challenge is particularly evident in studies of nocturnal hypoglycemia where physiological variability, the noise from CGM sensors, and the lack of concurrent behavioral data are all working against precise detection [15, 31, 61]. Research on hypoglycemia risk prediction emphasizes that the pursuit of maximum sensitivity often results in the loss of specificity [23, 45, 62]. Hence, the best clinical use of the model may still depend on patient-specific risk tolerance and context for model threshold adjustment rather than merely considering accuracy metrics.

In this review, a major limitation was pointed out, and it was the dataset homogeneity, which resulted in a lack of generalization. The use of high-frequency CGM sampling, well-annotated meals, or synthetic environments with reduced noise contributed a lot to the performance of the models that were assessed on the OhioT1DM [54], UVA/Padova simulations [53, 91], and DirecNet data [43], among others. On the other hand, real-world free-living datasets frequently had lower performance because of a lack of context, irregular meal reporting, and sensor inaccuracies [74, 79, 97].

A considerable part of the research depended on small in-house datasets (e.g.,$n<20$) or simulated environments. The studies that tested the models on the external cohorts

mostly reported performance drops, which underlined the transferability of models across different patient populations as a difficult issue. In addition, there were some socio-economic biases in the data collection; whereas some studies made good use of non-invasive health checkups in underdeveloped regions to develop cost-effective models, the majority of high-precision DL research is still highly dependent on the data-rich environments that are typical of the developed nations. Overcoming this challenge is contingent on the creation of large-scale, diverse, and publicly available datasets.

The predictive models aim at the realization of a fully automated AP system. The review reported that AI-integrated PID controllers and closed-loop systems do well in the management of post-prandial spikes and the maintenance of glucose stability. Nevertheless, the dependability of these systems in the real-world where situations like missed boluses, sensor calibration errors, and unannounced meals occur, still needs to be validated.

The "clinical accuracy" measurement (e.g., Clarke Error Grid) sometimes gives a clearer picture of a model's practical value than statistical error RMSE; among the models developed by Amar et al., which resulted in 99.3% clinical safety for short-term predictions, set the standard for further regulatory approval [6]. Although deep learning methods made their way to the top, the simpler ML methods are still fighting at the same level only if the conditions of the prediction horizon or the features structure are very well dealt with they are still to be considered for real-world deployment, where interpretability and computational efficiency are crucial.

To summarize, AI techniques for BGL prediction have reached a level of maturity that supports clinical decision making, but the field must now move from purely algorithmic improvements to robust, real-world validation. Multimodal and diverse datasets, improved generalization via external validation, models sharing predictive performance with interpretability, and personalized and context-aware prediction frameworks should be the focus of future research. These directions are necessary to prevent AI-based BGL prediction models from being stuck in the lab and instead transitioning to the clinic as trustworthy and clinical use, allowing equal and safe diabetes management over a vast range of global populations.

Conclusions

This review provides a comprehensive analysis of AI applications in predicting BGLs among diabetic patients. With the inclusion of 57 studies, it is clear that AI techniques have made significant progress, offering improved accuracy, real-time prediction, and personalized interventions. Deep learning models such as LSTM, CNN, and their hybrid forms consistently demonstrated strong predictive power, especially in short-term forecasts. Interpretable models like DTs and SHAP-enhanced LSTMs were also valuable in clinical settings due to their transparency. Nonetheless, key challenges persist. Many models depend on in-silico or small datasets, limiting generalizability. Non-invasive and wearable sensor integrations remain underdeveloped and under-tested in real-world environments. Moreover, while some approaches offer excellent accuracy, others trade off interpretability and clinical usability for performance.

To move toward practical clinical application, future studies should prioritize real-world validations, integration of multimodal physiological and lifestyle data, and user-centered design that includes feedback from healthcare professionals and patients. Combining the strengths of interpretable models, robust deep learning, and physiological insights will be essential to building safe, effective, and scalable AI solutions for glucose management in diabetes.

References

[1] Ahmed A., Aziz S., Abd-Alrazaq A., Farooq F., and et al., "The Effectiveness of Wearable Devices Using Artificial Intelligence for Blood Glucose Level Forecasting or Prediction: Systematic Review," *Journal of Medical Internet Research*, vol. 25, pp. 1-16, 2023. https://doi.org/10.2196/40259

[2] Alfian G., Syafrudin M., Anshari M., Benes F., and et al., "Blood Glucose Prediction Model for Type 1 Diabetes Based on Artificial Neural Network with Time-Domain Features," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1586-1599, 2020. https://doi.org/10.1016/j.bbe.2020.10.004

[3] Alhmiedat T and Alotaibi M., "The Investigation of Employing Supervised Machine Learning Models to Predict Type 2 Diabetes Among Adults," *KSII Transactions on Internet and Information Systems*, vol. 16, no. 9, pp. 2904-2926, 2022. DOI:10.3837/tiis.2022.09.004

[4] Alkanhel R., Saleh H., Elaraby A., Alharbi S., and et al., "Hybrid CNN-GRU Model for Real-Time Blood Glucose Forecasting: Enhancing IoT-Based Diabetes Management with AI," *Sensors*, vol. 24, no. 23, pp. 1-16, 2024. https://doi.org/10.3390/s24237670

[5] Alzboon M., Al-Batah M., Alqaraleh M., Abuashour A., and Hamadah Bader A., "Early Diagnosis of Diabetes: A Comparison of Machine Learning Methods," *International Journal of Online and Biomedical Engineering*, vol. 19, no. 15, pp. 144-165, 2023. https://online-journals.org/index.php/i-joe/article/view/42417

[6] Amar Y., Shilo S., Oron T., Amar E., and et al., "Clinically Accurate Prediction of Glucose Levels in Patients with Type 1 Diabetes," *Diabetes*

Technology and Therapeutics, vol. 22, no. 8, pp. 562-569, 2020. DOI: 10.1089/dia.2019.0435

[7] Anjum M., Saher R., and Saeed M., "Optimizing Type 2 Diabetes Management: AI-Enhanced Time Series Analysis of Continuous Glucose Monitoring Data for Personalized Dietary Intervention," *PeerJ Computer Science*, vol. 10, pp. 1-28, 2024. DOI: 10.7717/peerj-cs.1971

[8] Annuzzi G., Apicella A., Arpaia P., Bozzetto L.,and et al., "Impact of Nutritional Factors in Blood Glucose Prediction in Type 1 Diabetes Through Machine Learning," *IEEE Access*, vol. 11, pp. 17104-17115, 2023. DOI: 10.1109/ACCESS.2023.3244712

[9] Annuzzi G., Bozzetto L., Cataldo A., Criscuolo S., and Pesola M., "Predicting and Monitoring Blood Glucose Through Nutritional Factors in Type 1 Diabetes by Artificial Neural Networks," *Acta Imeko*, vol. 12, no. 2, pp. 1-7, 2023. DOI: https://doi.org/10.21014/actaimeko.v12i2.1453

[10] Aouragh A., Bahaj M., and Toufik F., "Diabetes Prediction: Optimization of Machine Learning through Feature Selection and Dimensionality Reduction," *International Journal of Online and Biomedical Engineering*, vol. 20, no. 8, pp. 100-114, 2024. DOI: https://doi.org/10.3991/ijoe.v20i08.47765

[11] Askari M., Rashid M., Sun X., Sevil M., and et al., "Meal and Physical Activity Detection from Free-living Data for Discovering Disturbance Patterns to Glucose Levels in People with Diabetes," *BioMedInformatics*, vol. 2, no. 2, pp. 297-317, 2022. DOI: 10.3390/biomedinformatics2020019

[12] Basile I and Sannino G., "Blood Glucose Level Prediction in Type 1 Diabetes: A Comparative Analysis of Interpretable Artificial Intelligence Approaches," *Results in Engineering*, vol. 25, no. 103681, pp. 1-10, 2025. https://doi.org/10.1016/j.rineng.2024.103681

[13] Beck R., "Closing in on Closed-Loop Systems for Type 2 Diabetes," *Nature Medicine*, vol. 29, pp. 33-34, 2023. https://doi.org/10.1038/s41591-022-02127-0

[14] Bertachi A., Biagi L., Contreras I., Luo N., and Vehi J., "Prediction of Blood Glucose Levels and Nocturnal Hypoglycemia Using Physiological Models and Artificial Neural Networks," *in Proceedings of the 3rd International Workshop on Knowledge Discovery in Healthcare Data, Stockholm*, Stockholm, pp. 1-6, 2018. https://api.semanticscholar.org/CorpusID:51951568

[15] Bertachi A., Vinals C., Biagi L., Contreras I., and et al., "Prediction of Nocturnal Hypoglycemia in Adults with Type 1 Diabetes Under Multiple Daily Injections Using Continuous Glucose Monitoring and Physical Activity Monitor," *Sensors*, vol. 20, no. 6, pp. 1-11, 2020. DOI: 10.3390/s20061705

[16] Bunescu R., Struble N., Marling C., Shubrook J., and Schwartz F., "Blood Glucose Level Prediction Using Physiological Models and Support Vector Regression," *in Proceedings of the IEEE 12th International Conference on Machine Learning and Applications*, Florida, Vol. 1, pp. 135-140, 2013. DOI: 10.1109/ICMLA.2013.30

[17] Cameron F., Bequette B., Wilson D., Buckingham B., Lee H., and Niemeyer G., "A Closed-Loop Artificial Pancreas Based on Risk Management," *Journal of Diabetes Science and Technology*, vol. 5, no. 2, pp. 368-379, 2011. DOI: 10.1177/193229681100500226

[18] Cloete L., "Diabetes Mellitus: An Overview of the Types, Symptoms, Complications and Management," *Nurs Stand*, vol. 37, no. 1, pp. 61-66, 2021. DOI: 10.7748/ns.2021.e11709

[19] Contreras I and Vehi J., "Artificial Intelligence for Diabetes Management and Decision Support: Literature Review," *Journal of Medical Internet Research*, vol. 20, no. 5, pp. 1-37, 2018. DOI: 10.2196/10775

[20] D'Antoni F., Merone M., Piemonte V., Iannello G., and Soda P., "Auto-Regressive Time Delayed Jump Neural Network for Blood Glucose Levels Forecasting," *Knowledge-Based Systems*, vol. 203, pp. 106134, 2020. https://doi.org/10.1016/j.knosys. 2020.106134

[21] Daniels J., Herrero P., and Georgiou P., "A Multitask Learning Approach to Personalized Blood Glucose Prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 436-445, 2022. https://doi.org/10.1109/JBHI.2021.3100558

[22] Dasilva J., "Diabetes," https://www.kaggle.com/johndasilva/diabetes Last Visited, 2025.

[23] Dave D., DeSalvo D., Haridas B., McKay S., and et al., "Feature-Based Machine Learning Model for Real-Time Hypoglycemia Prediction," *Journal of Diabetes Science and Technology*, vol. 15, no. 4, pp. 8420-855, 2021. DOI: 10.1177/1932296820922622

[24] Del Favero, S., Facchinetti A., and Cobelli C., "A Glucose-Specific Metric to Assess Predictors and Identify Models," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1281-1290, 2012. DOI: 10.1109/TBME.2012.2185234

[25] Dubosson F., Ranvier J., Bromuri S., Calbimonte J., and et al., "The Open D1NAMO Dataset: A Multi-Modal Dataset for Research on Non-Invasive Type 1 Diabetes Management," *Informatics in Medicine Unlocked*, vol. 13, pp. 92-100, 2018. https://doi.org/10.1016/j.imu.2018.09.003

[26] Ekeland A., Bowes A., and Flottorp S., "Methodologies for Assessing Telemedicine: A Systematic Review of Reviews," *International*

Journal of Medical Informatics, vol. 81, no. 1, pp. 1-11, 2012. DOI: 10.1016/j.ijmedinf.2011.10.009

[27] Elhadd T., Mall R., Bashir M., Palotti J., and et al., "Artificial Intelligence (AI) Based Machine Learning Models Predict Glucose Variability and Hypoglycaemia Risk in Patients with Type 2 Diabetes on a Multiple Drug Regimen Who Fast During Ramadan (The PROFAST-IT Ramadan Study)," *Diabetes Research and Clinical Practice*, vol. 169, pp. 1-9, 2020. DOI: 10.1016/j.diabres.2020.108388

[28] Ellahham S., "Artificial Intelligence: The Future for Diabetes Care," *American Journal of Medicine*, vol. 133, no. 8, pp. 895-900, 2020. DOI: 10.1016/j.amjmed.2020.03.033

[29] Eskaf K., "Advanced AI Techniques for Real-Time Blood Glucose Prediction in Diabetics: A Study Using Deep Learning and Genetic Algorithms," *Midocean Journal for Research and Scientific Studies*, vol. 1, no. 1, pp. 123-135, 2024. DOI: https://doi.org/10.71335/5y36na30

[30] Estremera E., Cabrera A., Beneyto A., and Vehi J., "A simulator with realistic and Challenging Scenarios for Virtual T1D Patients Undergoing CSII and MDI Therapy," *Journal of Biomedical Informatics*, vol. 132, art no. 104141, pp. 1-13, 2022. DOI: 10.1016/j.jbi.2022.104141

[31] Faccioli S., Prendin F., Facchinetti A., Sparacino G., and Del Favero S., "Combined Use of Glucose-Specific Model Identification and Alarm Strategy Based On Prediction-Funnel to Improve Online Forecasting of Hypoglycemic Events," *Journal of Diabetes Science and Technology*, vol. 17, no. 5, pp. 1295-1303, 2023. DOI: 10.1177/19322968221093665

[32] Floris L., "Type 1 Diabetes Blood Glucose Prediction," *Kaggle*, https://www.kaggle.com/datasets/lacofloris/type-1-diabetes-blood-glucose-prediction, Last Visited, 2025.

[33] Georga E., Protopappas V., Ardigo D., Marina M., and et al., "Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, pp. 71-81, 2013. DOI: 10.1109/TITB.2012.2219876

[34] Georga E., Protopappas V., Polyzos D., and Fotiadis D., "A Predictive Model of Subcutaneous Glucose Concentration in Type 1 Diabetes Based on Random Forests," *in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, California, 2889-2892, 2012. DOI: 10.1109/EMBC.2012.6346565

[35] Guemes A., Cappon G., Hernandez B., Reddy M., and et al., "Predicting Quality of Overnight Glycaemic Control in Type 1 Diabetes Using Binary Classifiers," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, pp. 1439-1446, 2020. DOI: 10.1109/JBHI.2019.2938305

[36] Haidar A., Wilinska M., Graveston J., and Hovorka R., "Stochastic Virtual Population of Subjects with Type 1 Diabetes for The Assessment of Closed-Loop Glucose Controllers," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 12, pp. 3524-3533, 2013. DOI: 10.1109/TBME.2013.2272736

[37] Hameed E., Joshi H., and Kadhim Q., "Advancements in Artificial Intelligence Techniques for Diabetes Prediction: A Comprehensive Literature Review," *Journal of Robotics and Control*, vol. 6, no. 1, pp. 345-365, 2025. https://doi.org/10.18196/jrc.v6i1.22258

[38] Heinemann L. and Freckmann G., "CGM Versus FGM; or, Continuous Glucose Monitoring is Not Flash Glucose Monitoring," *Journal of Diabetes Science and Technology*, vol. 9, no. 5, pp. 947-950, 2015. DOI: 10.1177/1932296815603528

[39] Institute for Health Metrics and Evaluation, Global Diabetes Cases To Soar From 529 Million to 1.3 Billion by 2050, https://www.healthdata.org/news-events/newsroom/news-releases/global-diabetes-cases-soar-529-million-13-billion-2050, Last Visited, 2025

[40] International Diabetes Federation, About Diabetes, https://worlddiabetesday.org/about-wdd/, Last Visited,2025

[41] Jacobs P., Herrero P., Facchinetti A., Vehi J., and et al., "Artificial Intelligence and Machine Learning for Improving Glycemic Control in Diabetes: Best Practices, Pitfalls, and Opportunities," *IEEE Reviews in Biomedical Engineering*, vol. 17, pp. 19-41, 2024. DOI: 10.1109/RBME.2023.3331297

[42] Jaeb Center for Health Research, "Listing of Available T1D Data Sets," https://public.jaeb.org/datasets/diabetes, Last Visited, 2025.

[43] Jaeb Center for Health Research, DirecNet Dataset, "Evaluation of Counter-Regulatory Hormone Responses during Hypoglycemia and the Accuracy of Continuous Glucose Monitors in Children with T1DM," ClinicalTrials.gov, NCT00607139, https://clinicaltrials.gov/study/NCT00607139, Last Visited, 2025.

[44] Jahangir Z., Shah Y., Qureshi S., Qureshi H., and etal., "From Data to Decisions: The AI Revolution in Diabetes Care," *International Journal of Membrane Science and Technology*, vol. 10, no. 5, pp. 1162-1179, 2023. https://www.researchgate.net/publication/383359865_From_Data_to_Decisions_The_AI_Revolution_in_Diabetes_Care

[45] Jensen M., Dethlefsen C., Vestergaard P., and Hejlesen, O., "Prediction of Nocturnal Hypoglycemia from Continuous Glucose Monitoring Data in People with Type 1 Diabetes: A Proof-of-Concept Study," *Journal of Diabetes Science and Technology*, vol. 14, no. 2, pp. 250-256, 2020. DOI: 10.1177/1932296819868727

[46] Keuning W., Medtronic MiniMed 640G and MiniMed 670G, https://github.com/ictinnovaties-zorg/dataset-diabetes-adolescents-time-series-with-heart-rate, GitHub, Last Visited, 2025.

[47] Kushner T., Breton M., and Sankaranarayanan S., "Multi-Hour Blood Glucose Prediction in Type 1 Diabetes: A Patient-Specific Approach Using Shallow Neural Network Models," *Diabetes Technology and Therapeutics*, vol. 22, no. 12, pp. 883-891, 2020. DOI: 10.1089/dia.2020.0061

[48] Li K., Daniel J., Liu C., Herrero P., and Georgiou P., "Convolutional recurrent neural networks for glucose prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 603-613, 2020. DOI: 10.1109/JBHI.2019.2908488

[49] Li K., Liu C., Zhu T., Herrero P., and Georgiou P., "GluNet: A Deep Learning Framework for Accurate Glucose Forecasting," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 414-423, 2020. DOI: 10.1109/JBHI.2019.2931842

[50] Linares-Espinos E., Hernandez V., Dominguez-Escrig J., Fernandez-Pello S., and et al., "Methodology of a Systematic Review," *Actas Urologicas Espanolas (English Edition)*, vol. 42, no. 8, pp. 499-506, 2018. DOI: 10.1016/j.acuro.2018.01.010

[51] Lindner N., Kuwabara A., and Holt T., "Non-Invasive and Minimally Invasive Glucose Monitoring Devices: A Systematic Review and Meta-Analysis on Diagnostic Accuracy of Hypoglycaemia Detection," *Systematic Reviews*, vol. 10, no. 1, pp. 1-14, 2021. DOI: 10.1186/s13643-021-01644-2

[52] Liu K., Li L., Ma Y., Jiang J., and et al., "Machine Learning Models for Blood Glucose Level Prediction in Patients with Diabetes Mellitus: Systematic Review and Network Meta-Analysis," *JMIR Medical Informatics*, vol. 11, no. e47833, pp. 1-30, 2023. DOI: 10.2196/47833

[53] Man C., Micheletto F., Lv D., Breton M., and et al., "The UVA/PADOVA Type 1 Diabetes Simulator: New Features," *Journal of Diabetes Science and Technology*, vol. 8, no. 1, pp. 26-34, 2014. DOI: 10.1177/1932296813514502

[54] Marling C. and Bunescu R., "The OhioT1DM Dataset for Blood Glucose Level Prediction: Update 2020," *CEUR Workshop Proceedings*, vol. 2675, pp. 71-74, 2020. https://webpages.charlotte.edu/rbunescu/data/ohiot1dm/OhioT1DM-dataset.html

[55] Martinez-Delgado L., Munoz-Organero M., and Queipo-Alvarez P., "Using Absorption Models for Insulin and Carbohydrates and Deep Leaning to Improve Glucose Level Predictions," *Sensors*, vol. 21, no. 16, pp. 1-29, 2021. https://doi.org/10.3390/s21165273

[56] Mayer-Davis E., Lawrence J., Dabelea D., Divers J., and et al "Incidence Trends of Type 1 and Type 2 Diabetes Among Youths, 2002-2012," *New England Journal of Medicine*, vol. 376, no. 15, pp. 1419-1429, 2017 DOI: 10.1056/NEJMoa1610187

[57] McShinsky R and Marshall B., "Comparison of Forecasting Algorithms for Type 1 Diabetic Glucose Prediction on 30 and 60-Minute Prediction Horizons," *in Proceedings of the 5th International Workshop on Knowledge Discovery in Healthcare Data (KDH@ECAI)*, vol, 2675, pp. 12-18, 2020. https://ceur-ws.org/Vol-2675/paper2.pdf

[58] Medanki S., Dommati N., Bodapati H., Katru V., and et at., "Artificial Intelligence Powered Glucose Monitoring and Controlling System: Pumping Module," *World Journal of Experimental Medicine*, vol. 14, no. 1, pp. 1-13, 2024. DOI: 10.5493/wjem.v14.i1.87916

[59] Mohebbi A., Johansen A., Hansen N., Christensen P., and et al., "Short Term Blood Glucose Prediction based on Continuous Glucose Monitoring Data," *in Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Quebec, pp. 5140-5145, 2020. https://doi.org/10.1109/EMBC44109.2020.9176695

[60] Mosquera-Lopez C and Jacobs P., "Incorporating Glucose Variability into Glucose Forecasting Accuracy Assessment Using the New Glucose Variability Impact Index and the Prediction Consistency Index: An LSTM Case Example," *Journal of Diabetes Science and Technology*, vol. 16, no. 1, pp. 7-18, 2022. DOI: 10.1177/19322968211042621

[61] Mosquera-Lopez C., Dodier R., Tyler N., Wilson L., and et al., "Predicting and Preventing Nocturnal Hypoglycemia in Type 1 Diabetes Using Big Data Analytics and Decision Theoretic Analysis," *Diabetes Technology and Therapeutics*, vol. 22, no. 11, pp. 801-811, 2020. DOI: 10.1089/dia.2019.0458

[62] Oviedo S., Contreras I., Quiros C., Gimenez M., and et al., "Risk-Based Postprandial Hypoglycemia Forecasting Using Supervised Learning," *International Journal of Medical Informatics*, vol. 126, pp. 1-8, 2019. DOI: 10.1016/j.ijmedinf.2019.03.008

[63] Pappada S., Cameron B., Rosman P., Bourey R., and et al., "Neural Network-Based Real-Time Prediction of Glucose in Patients with Insulin-

Dependent Diabetes," *Diabetes Technology and Therapeutics*, vol. 13, no. 2, pp. 135-141, 2011. DOI: 10.1089/dia.2010.0104

[64] Parcerisas A., Contreras I., Delecourt A., Bertachi A., and et al., "A Machine Learning Approach to Minimize Nocturnal Hypoglycemic Events in Type 1 Diabetic Patients Under Multiple Doses of Insulin," *Sensors*, vol. 22, no. 4, pp. 1-14, 2022. DOI: 10.3390/s22041665

[65] Perez-Gandia C., Facchinetti A., Sparacino G., Cobelli C., and et al., "Artificial Neural Network Algorithm for Online Glucose Prediction from Continuous Glucose Monitoring," *Diabetes Technology and Therapeutics*, vol. 12, no. 1, pp. 81-88, 2010. DOI: 10.1089/dia.2009.0076

[66] Plis K., Bunescu R.., Marling C., Shubrook J., and Schwartz F., "A Machine Learning Approach to Predicting Blood Glucose Levels for Diabetes Management," *in Proceedings of the AAAI Workshop: Modern Artificial Intelligence for Health Analytics*, Quebec, pp. 35-39, 2014. https://api.semanticscholar.org/CorpusID:17274044

[67] Pohanka M., "Quartz Crystal Microbalance (QCM) Sensing Materials in Biosensors Development," *International Journal of Electrochemical Science*, vol. 16, no. 12, pp.1-10, 2021. https://doi.org/10.20964/2021.12.15

[68] Prendin F., Del Favero S., Vettoretti M., Sparacino G., and Facchinetti A., "Forecasting of Glucose Levels and Hypoglycemic Events: Head-to-Head Comparison of Linear and Nonlinear Data-Driven Algorithms Based on Continuous Glucose Monitoring Data Only," *Sensors*, vol. 21, no. 5, pp. 1-21, 2021. https://doi.org/10.3390/s21051647

[69] Prendin F., Pavan J., Cappon G., Del Favero S., and et al, "The Importance of Interpreting Machine Learning Models for Blood Glucose Prediction in Diabetes: An Analysis Using SHAP," *Scientific reports*, vol. 13, no. 1, pp. 1-14, 2023. DOI: 10.1038/s41598-023-44155-x

[70] Ramachandran A., "Know the Signs and Symptoms of Diabetes," *Indian Journal of Medical Research*, vol. 140, no. 5, pp. 579-581, 2014. https://www.researchgate.net/publication/270825871_Know_the_signs_and_symptoms_of_diabetes

[71] Ramalingaswamy Cheruku D.and Kuppili V., "An Optimized and Efficient Radial Basis Neural Network Using Cluster Validity Index for Diabetes Classification," *The International Arab Journal of Information Technology*, vol. 16, no. 5, pp. 816-826, 2019.

[72] Rashid M., Askari M., Chen C., Liang Y., and et al., "Artificial Intelligence Algorithms for Treatment of Diabetes," *Algorithms*, vol. 15, no. 9, pp. 1-20, 2022. DOI: 10.3390/a15090299

[73] Rashid M., Samadi S., Sevil M., Hajizadeh I., and et al., "Simulation Software for Assessment of Nonlinear and Adaptive Multivariable Control Algorithms: Glucose-Insulin Dynamics in Type 1 Diabetes," *Computers and Chemical Engineering*, vol. 130, 2019. DOI: 10.1016/j.compchemeng.2019.106565

[74] Reddy R., Resalat N., Wilson L., Castle J., and et al., "Prediction of Hypoglycemia During Aerobic Exercise in Adults with Type 1 Diabetes," *Journal of Diabetes Science and Technology*, vol. 13, no. 5, pp. 919-927, 2019. DOI: 10.1177/1932296818823792

[75] Resalat N., El Youssef J., Tyler N., Castle J., and Jacobs P., "A Statistical Virtual Patient Population for the Glucoregulatory System in Type 1 Diabetes with Integrated Exercise Model," *Plos One*, vol. 14, no. 7, pp. 1-11, 2019. DOI: 10.1371/journal.pone.0217301

[76] Rodriguez Leon C., Banos O., Fernandez Mora O., Martinez Bedmar A., and et al., "Prediction of Blood Glucose Levels in Patients with Type 1 Diabetes via LSTM Neural Networks," *in proceeding of the 17th International Work-Conference on Artificial Neural Networks*, Ponta Delgada, pp. 563-573, 2023. DOI: https://doi.org/10.1007/978-3-031-43085-5_45

[77] Rodriguez-Leon C., Aviles-Perez M., Banos O., Quesada-Charneco M., and et al, "T1DiabetesGranada: a longitudinal multi-modal dataset of type 1 diabetes mellitus," *Scientific Data*, vol. 10, no. 1, pp. 1-11, 2023. https://doi.org/10.1038/s41597-023-02737-4

[78] Ruan Y., Bellot A., Moysova Z., Tan G., and et al., "Predicting the Risk of Inpatient Hypoglycemia with Machine Learning Using Electronic Health Records," *Diabetes Care*, vol. 43, no. 7, pp. 1504-1511, 2020. DOI: 10.2337/dc19-1743

[79] Sampa M., Biswas T., Rahman M., Aziz N., and et al., "A Machine Learning Web App to Predict Diabetic Blood Glucose Based on a Basic Noninvasive Health Checkup, Sociodemographic Characteristics, and Dietary Information: Case Study," *JMIR diabetes*, vol. 8, no. 5, pp. 1-11, 2023. DOI: 10.2196/49113

[80] Sampath S., Tkachenko P., Renard E., and Pereverzev S., "Glycemic Control Indices and Their Aggregation in the Prediction of Nocturnal Hypoglycemia from Intermittent Blood Glucose Measurements," *Journal of Diabetes Science and Technology*, vol. 10, no. 6, pp. 1245-1250, 2016. DOI: 10.1177/1932296816670400

[81] Seo W., Lee Y., Lee S., Jin S., and Park S., "A Machine Learning Approach to Minimize Nocturnal Hypoglycemic Events in Type 1 Diabetic Patients under Multiple Doses of Insulin," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1-13, 2019. DOI:

10.1186/s12911-019-0943-4

[82] Smith J., Everhart J., Dickson W., Knowler W., and Johannes R., "Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus," *in Proceedings of the Symposium on Computer Applications and Medical Care, IEEE Computer Society Press*, Washington, pp. 261-265, 1988. https://pmc.ncbi.nlm.nih.gov/articles/PMC2245318/

[83] Srinivasu P., Ahmed S., Hassaballah M., and Almusallam N., "An Explainable Artificial Intelligence Software System for Predicting Diabetes," *Heliyon*, vol. 10, no. 16, pp. 1-19, 2024. https://doi.org/10.1016/j.heliyon.2024.e36112

[84] Stuart K., Adderley N., Marshall T., Rayman G., and et al., "Predicting Inpatient Hypoglycaemia in Hospitalized Patients with Diabetes: A Retrospective Analysis of 9584 Admissions with Diabetes," *Diabetic Medicine*, vol. 34, no. 10, pp. 1385-1391, 2017. DOI: 10.1111/dme.13409

[85] Subramanian A., Adap S., Chawale S., Singh S., and Sudhakaran P., "Non Invasive Glucose Measurement Using Raman Spectroscopy," *International Research Journal of Engineering and Technology*, vol. 4, no. 03, pp. 2872-2875, 2017. https://www.irjet.net/archives/V4/i3/IRJET-V4I3722.pdf

[86] Sudharsan B., Peeples M., and Shomali M., "Hypoglycemia Prediction Using Machine Learning Models for Patients with Type 2 Diabetes," *Journal of Diabetes Science and Technology*, vol. 9, no. 1, pp. 86-90, 2015. DOI: 10.1177/1932296814554260

[87] Tang L., Chang S., Chen C., and Liu J., "Non-Invasive Blood Glucose Monitoring Technology: A Review," *Sensors*, vol. 20, no. 23, pp. 1-32, 2020. DOI: 10.3390/s20236925

[88] Tidepool, "Tidepool Big Data Donation Data Set," https://support.tidepool.org/hc/en-us/articles/360044350552-Exporting-Tidepool-data, Last Visited, 2025.

[89] Wang J., "Electrochemical Glucose Biosensors," *Chemical Reviews*, vol. 108, no. 2, pp. 814-825, 2008. https://doi.org/10.1021/cr068123a

[90] Wang Y., "A Comparison of Machine Learning Algorithms in Blood Glucose Prediction for People with Type 1 Diabetes," *in Proceedings of the 2nd International Symposium on Artificial Intelligence for Medicine Sciences*, Beijing, pp. 351-360, 2021. https://doi.org/10.1145/3500931.3500993

[91] Wilinska M., Chassin L., Acerini C., Allen J., and et al., "Simulation Environment to Evaluate Closed-Loop Insulin Delivery Systems in Type 1 Diabetes," Journal of Diabetes Science and Technology, vol. 4, no. 1, pp. 132-144, 2010. DOI: 10.1177/193229681000400117

[92] Yismaw M., Tafere C., Tefera B., Demsie D., and et al., "Artificial Intelligence Based Predictive Tools for Identifying Type 2 Diabetes Patients at High Risk of Treatment Non-Adherence: A Systematic Review," *International Journal of Medical Informatics*, vol. 198, pp. 105858, 2025. DOI: 10.1016/j.ijmedinf.2025.105858

[93] Zecchin C., Facchinetti A., Sparacino G., and Cobelli C., "Jump Neural Network for Real-Time Prediction of Glucose Concentration," *Methods in Molecular Biology*, vol. 1260, pp. 245-259, 2015. DOI: 10.1007/978-1-4939-2239-0_15

[94] Zecchin C., Facchinetti A., Sparacino G., De Nicolao G., and Cobelli C., "Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1550-1560, 2012. DOI: 10.1109/TBME.2012.2188893

[95] Zhang M., Flores K., and Tran H., "Deep Learning and Regression Approaches to Forecasting Blood Glucose Levels for Type 1 Diabetes," *Biomedical Signal Processing and Control*, vol. 69, pp. 102923, 2021. DOI: 10.1016/j.bspc.2021.102923

[96] Zhu T., Li K., Chen J., Herrero P., and Georgiou P., "Dilated Recurrent Neural Networks for Glucose Forecasting in Type 1 Diabetes," *Journal of Healthcare Informatics Research*, vol. 4, no. 3, pp. 308-324, 2020. DOI: 10.1007/s41666-020-00068-2

[97] Zhu T., Uduku C., Li K., Herrero P., and Georgiou P., "Enhancing self-management in type 1 diabetes with wearables and deep learning," *NPJ Digital Medicine*, vol. 5, no. 78, pp. 1-11, 2022. DOI: 10.1038/s41746-022-00626-5

Mohammed Alotaibi received his Ph.D. in Health Informatics from Kingston University in 2015. He has served in various academic and administrative roles at the University of Tabuk. He currently leads the Data Management Office at the University of Hafr Al Batin and serves as a Data Management Consultant at Sulaiman AlRajhi University. His professional interests include Health Informatics, Data Management, and Digital Health.

Osama Alia received his Ph.D. in Computer Science with a specialization in Artificial Intelligence from Universiti Sains Malaysia in 2011. He has been with the University of Tabuk since 2011, where he has served in various academic and administrative roles. His research interests include Artificial Intelligence, Machine Learning, Deep Learning, Computer Vision and Data Analytics.

Table 1. Comparative overview of existing research on artificial intelligence methodologies for predicting blood glucose levels.

Ref.	AI Technique	Results	Performance Metrics (RMSE, AUC, Sensitivity, etc.)	Implications	Year
[9]	ANN	Predicted BGLs 60 minutes after meals using different nutritional factors such as carbohydrates, proteins, lipids, fibers, and meal energy intake.	Predictions improved significantly with patient-specific data.	Shows the importance of considering nutritional factors for the precise BGL prediction.	2023
[8]	ANN	Predicted BGLs at intervals (15, 30, 45, 60 minutes) after meals and showed that nutritional factors significantly influence short- and mid-term BGL predictions.	Models specific to patients showed better results than generic models.	Supports the integration of diverse nutritional data into AI systems to better manage diabetes.	2023
[16]	SVR	Modeled patient-specific BGLs with the use of physiological models. Showed better performance than expert predictions.	Outperformed diabetes experts in predicting trends about blood glucose.	Shows the effectiveness of combining physiological models with AI for personalized diabetes care.	2013
[65]	ANN	Reliable short-term BGL predictions with the use of CGM data. Outperformed traditional autoregressive models.	RMSE: 10 mg/dl (15 min), 18 mg/dl (30 min), 27 mg/dl (45 min); the prediction delay was minimal.	Offers a practical solution for glucose prediction in real-time for preventing hypo/hyperglycemia.	2010
[7]	XGBoost, SARIMA, Prophet	Monitored trends in glucose levels with the use of CGM data, which helped personalized dietary adjustments decrease HbA1c.	High for normalizing blood glucose changes; and detailed accuracy metrics were not provided.	Improved glucose control and healthier lifestyle for T2D patients by personalizing their interventions on the basis of CGM data.	2024
[27]	XGBoost	High performance for predicting normal glucose (95.2%) and hyperglycemia (82.6%), limited for hypoglycemia (27.9%).	$R^2 = 0.836$, MAE = 17.47.	Helps manage glucose variability during fasting periods like Ramadan while considering demographics, BMI, and medication use.	2020
[66]	SVR	Anticipated 25% of hypoglycemic events 30 minutes in advance; precision for hypoglycemia was 42%.	Moderate accuracy of prediction for hypoglycemia; there were false alarms in near-hypoglycemic regions.	Allows time for preventative action but needs improvement for fewer false positives.	2014
[58]	PID Controller integrated with AI decision unit	Maintained glucose levels over a longer duration by adjusting insulin dosage dynamically post-meal.	Effective for dynamic insulin control; no specific metrics were provided.	Offers automated insulin dosing through closed-loop systems, which improves long-term glucose stability and decreases post-meal spikes.	2024
[72]	Novel ML algorithms, automated insulin delivery models	Demonstrated usefulness in analyzing EHR and wearable device data to improve insulin delivery systems.	Not quantified but showed improved decision-making for chronic disease management with the use of AI.	Supports the development of trustworthy, reliable, and effective AI tools for better treatment and outcomes in the management of diabetes.	2022
[12]	DT, Multi-series Regression	DT got a weighted F1 score of 0.87, outperforming other models in accurate glucose prediction.	Weighted F1 score of 0.87 for DT model.	Shows the potential for non-invasive glucose monitoring and the value of interpretable AI models in the improvement insulin therapy and patient adherence.	2025
[2]	ANN, Time-domain features	ANN outperformed other models with the lowest RMSE for 15-60 minute prediction horizons.	RMSE: 2.82, 6.31, 10.65, and 15.33 mg/dL for 15, 30, 45, and 60 min, respectively.	Time-domain features improved AI model performance and offered preventive alerts for hypo- and hyperglycemic events.	2020
[57]	OLS, LASSO, Ridge, Elastic Net, RBF, Poly, KNN, RF, Grad, XGB, Kalman, UKF, ARIMA, VAR, ANFIS, MLP	LASSO regression model worked better than other algorithms for both time horizons	Average RMSE of 20.58 mg/dL (1.14 mmol/L) on the 30-minute prediction horizon Average RMSE of 33.41 mg/dL (1.86 mmol/L) on the 60-minute prediction horizon	Demonstrates progress in BG prediction, it also highlights limitations and areas for improvement, particularly for long-term predictions and hypoglycemic event detection. Identified glucose level, bolus, meal, and exercise as the most relevant features for predicting BG levels.	2020
[95]	MLR, BRC, DCNN, Seq-to-Seq LSTM	Seq-to-Seq LSTM model was most effective in predicting BGLs within a 30-minute horizon. MLR model performed better for a prediction horizon of 60 minutes	Seq-to-Seq LSTM : RMSE of 15.46 mg/dL (0.86 mmol/L) MLR: RMSE of 21.59 mg/dL (1.2 mmol/L) with lower computational costs	Advanced predictive techniques especially DL models which can enhance the performance and safety of closed-loop insulin delivery systems by enabling more accurate glucose prediction. IT could lead to tighter glycemic control and reduce the risk of hypo- or hyperglycemic events in individuals with T1D	2021
[34]	RF Regression	RMSE 6.60, 8.15, 9.25, 10.83 mg/dL (15,30,60,120 min)	Low prediction errors at short horizons; gradual increase over time	Demonstrated feasibility of RF for glucose forecasting	2012
[33]	Support Vector Regression	RMSE ~5.21, 6.03, 7.14, 7.62 mg/dL (15–120 min)	Improved accuracy using multivariable inputs	Multivariable SVR improves short-term glucose forecasting	2013
[63]	ANN	RMSE 43.9 mg/dL; MAPE 22.1%; Clarke 92.3% in zones A/B	Moderate; tends to overestimate hypoglycemia	Feasible real-time BG prediction for insulin-dependent patients	2011
[93]	Jump NN	Feasible 30-min ahead prediction (no numeric metrics provided)	Sufficient for generating preventive alerts	Enables predictive alerts for hypoglycemia	2015

[94]	ANN	Improved short-term prediction accuracy with meal data	RMSE reduced by ~18%, from ~25 to ~20.5 mg/dL	Demonstrated value of integrating meal data for short-term predictions	2012
[47]	Shallow NN (patient-specific)	RMSE 28–43 mg/dL for 60–240 min; ≥93% predictions clinically A/B	Good over multi-hour horizons	Enables multi-hour forecasting; supports closed-loop therapy	2020
[45]	Classification model (feature-based)	AUC 0.79; Sens 75%, Spec 70%	Moderate night-hypo prediction accuracy	Proof-of-concept for nocturnal hypo alerting using CGM data	2020
[24]	Glucose-specific MSE metric	Penalizes clinically risky errors more than standard MSE	Not predictive (evaluation metric)	Better evaluation metric reflecting clinical glucose risk	2012
[31]	ARIMAX + gMSE + prediction funnel	Precision 65%, Recall 88%, F1 75%; 0.29 FP/day; 15 min lead	High recall for hypoglycemia forecasting	Combined strategy improves hypoglycemia prediction	2023
[17]	Model Predictive Control (EMPC)	92% BG in target range, mean ~140 mg/dL; BGRI 1.27 (56% reduction)	Effective BG control with low hypo risk	Risk-aware control better avoids hypoglycemia	2011
[68]	Linear (ARIMA/ARMA) vs Nonlinear (SVR, RF, ANN, LSTM)	Best RMSE ~21.5 mg/dL (nonlin), 22.1 (lin); Hypo prec ~63–64%, rec 69–82%	Comparable performance between linear and nonlinear models	Simple ARIMA nearly as good as ML for CGM-only forecasting	2021
[20]	AR time-delayed Jump Neural Network	Reported improved forecast performance (detailed metrics not provided)	Higher accuracy than standard approaches	Combines AR input with jump NN for enhanced BG prediction	2020
[6]	ANN (CEG-optimized)	99.3% Clarke A/B at 30 min; 95.8% at 60 min	Very high clinical accuracy	Optimizing for clinical zones yields safest predictions	2020
[81]	RF	AUC 0.966; Sens 89.6%, Spec 91.3%, F1 0.543 (30 min)	Very high accuracy for meal-time hypo prediction	RF effectively forecasts post-meal hypoglycemia	2019
[64]	ML (SVM, etc.)	30g CHO intervention reduces time <70 mg/dL by 35%; time in range +1.3%	Demonstrated significant risk reduction	ML-informed intervention (snack) can prevent nocturnal hypo	2022
[84]	LR	AUC 0.733	Moderate prediction accuracy	Electronic records can identify hypo risk inpatients	2017
[15]	SVM	Sens 78.8%, Spec 82.2%; >70% NH events avoidable	High predictive performance	Incorporating activity data significantly improves NH prediction	2020
[61]	SVR	Predicted 94.1% of NH events; AUC 0.86; reduced NH by 77% (in silico)	Very high prediction accuracy	SVR + decision theory can greatly reduce NH incidence	2020
[78]	XGBoost	AUC 0.96 (XGBoost) vs 0.75 (logistic)	Very high (XGBoost)	Advanced ML substantially outperforms LR	2020
[35]	Binary classification algorithms	AUC ~0.70	Moderate prediction accuracy	Daytime data can predict overnight glycemic outcome	2020
[62]	Supervised ML	Demonstrated improved risk assessment (details not specified)	Not quantified	Novel approach for meal-related hypo risk forecasting	2019
[14]	ANN + Physiological models	RMSE ~19–32 mg/dL (30/60 min); Hypo sens 66.7–100%	Reasonable for both BG and hypo	ANNs are feasible for continuous and event prediction	2018
[74]	DT (180/120 rule), RF	DT 79.6% accuracy; RF 86.7% accuracy	High	Models (simple or complex) aid avoidance of exercise hypo	2019
[86]	Probabilistic ML	Sens 92%, Spec 70% (24h horizon); Spec 90% (1h with medication data)	High sensitivity; specificity improved with medication context	ML can accurately predict hypo using sparse SMBG & meds	2015
[3]	LR, KNN, DT, GB, RF, SVM, Light GB, CatBoost, ANN and CNN.	Accuracy 98.95%	Very high classification accuracy	Enhanced Diabetes Prediction with Machine Learning	2022
[5]	LR, KNN, DT, GB, RF, SVM, Light GB, CatBoost, ANN and CNN.	LR and ANN were the best algorithms in predicting T2D in most evaluation criteria.	Very high classification accuracy	Enhanced Diabetes Prediction with Machine Learning	2023
[29]	DL, Genetic Algorithms, RL	Improved BGL prediction accuracy with the use of advanced AI methods. Genetic Algorithms outperformed artificial neural networks (ANNs). Accuracy was further improved by RL.	Genetic algorithms achieved the highest accuracy for short-term predictions (+30 minutes).	Shows the value of using advanced AI techniques for real-time diabetes management.	2024
[69]	SHAP, LSTM (p-LSTM and np-LSTM)	p-LSTM model learned physiological relationships and improved glycemic control when integrated into DSS.	Both LSTMs had similar prediction accuracies; however, p-LSTM performed better in therapeutic decision support.	Emphasizes the importance of model interpretability to ensure physiologically-sound predictions for safer management of diabetes.	2023
[90]	LSTM, SVM, RF, KNN, XGBoost, LR	Bi-LSTM achieved the best results for 30-minute predictions.	RMSE = 7.55 ± 0.19 mg/dL, R^2 = 0.96 for Bi-LSTM.	Shows the superior performance of Bi-LSTM in predicting glucose levels accurately, which supports its use for patient self-management.	2021
[11]	RNN model, LSTM networks	Predicted the occurrence of meals and exercise events based on CGM and insulin pump data.	Accuracy 94.72%, Weighted recall 94.72%, F1 scores up to 96.85%	Demonstrates that deep learning models can enhance automated insulin delivery systems by predicting disturbances caused by meals and physical activity. RNN-based models can improve personalized diabetes management, potentially reducing the risks of hypo- and hyperglycemia.	2022

[96]	DRNN	DRNN model outperformed traditional RNN models in terms of the precision and robustness of the prediction	RMSE: 18.9 mg/dL (1.05 mmol/L)	Significantly improve the accuracy of BG predictions for individuals with type T1D, it may support the close loop systems where the AI model can be integrated with CGM and insulin pump to optimize the insulin delivery.	2020
[76]	LSTM Neural Networks	successfully used LSTM neural networks to predict blood glucose levels at multiple future intervals (30, 60, 90, and 180 minutes). The model produced clinically reliable predictions across all horizons.	Achieved over 95% clinical accuracy according to Clarke Error Grid analysis	Demonstrates LSTM can accurately predict BG levels up to 180 minutes in advance. Highlighting their potential for improving early intervention and personalized management in T1D	2023
[97]	DL (multi-modal)	60-min RMSE 35.28±5.77 mg/dL; MCC 0.56 (hypo), 0.70 (hyper)	Moderate; wristband data reduced RMSE by ~2.25 mg/dL	Multi-modal deep learning improves self-management and monitoring	2022
[49]	Dilated CNN	Significant RMSE improvements (state-of-art)	Very high (best-reported results)	Advanced DL yields top-tier CGM forecasting	2020
[59]	RNN	Population RNN outperformed ARIMA up to 90 min horizon	Better accuracy than linear models	RNNs effective for CGM-only short-term predictions	2020
[21]	Multitask DL	RMSE 18.8–47.2 mg/dL (30–120 min); ≥93% Clarke A/B	High personalized accuracy	Transfer-learning improves model personalization	2022
[79]	Boosted DT Regression	Average deviation of 2.30 mg/dL from actual blood glucose levels and supports regular monitoring.	RMSE = 2.30.	Cost-effective solution for the prediction of BGLs in resource-constrained areas and even encourages self-monitoring and community health.	2023
[83]	Bi-LSTM, CNN	Effectively identified people with high glucose levels with the use of spectrogram image classification.	Not explicitly stated but shown to be effective on the basis of evolutionary metrics and confusion matrix evaluation.	The AI system offers transparent glucose monitoring, increasing patient understanding and improving trust in AI-based healthcare.	2024
[4]	CNN, GRU, Hybrid Model	Hybrid CNN-GRU model outperformed standalone CNN, GRU, and LSTM models in real-time blood glucose forecasting.	Better accuracy and timeliness in comparison with individual models (quantitative metrics not specified).	Shows the value of integrating deep learning models with IoT-based systems for efficient and real-time glucose management.	2024
[55]	RNN-LSTM	Seven experiments modifying the insulin and carbohydrate absorption function	RMSE of 9.2 mg/dL (0.510 mmol/L) over a prediction horizon of 60 minutes.	Demonstrates that ML models can achieve more accurate predictions of BG levels. This can help patients better manage their condition and avoid hypoglycemia or hyperglycemia episodes. Highlights the importance of pre-processing data using physiological absorption models for insulin and carbohydrates.	2021
[60]	LSTM	RMSE ~19.8 mg/dL (30 min), 33.2 (60 min); 99.6% (30m) and 97.6% (60m) predictions in A/B	High accuracy; outperformed other models	LSTM effective on large CGM dataset; suggests new variability metrics	2022
[48]	CNN-RNN	Simulated RMSE 9.38 (30m), 18.87 (60m); Real RMSE 21.07, 33.27	High; minimal latency (6ms)	Deep CNN+RNN yields accurate mobile glucose predictions	2020
[23]	RF (feature-engineered)	Sens >91%, Spec >90% for 30-60 min; nocturnal sens ~95%	Very high sensitivity and specificity	Feature selection yields timely hypo alerts	2021
[80]	Aggregation of glycemic indices	Sens 77%, Spec 83.4%, PPV 80.2%, NPV 80.6%	Good	Aggregated indices boost telemedicine NH prediction	2016

Table 2. Overview of datasets utilized in the examined studies, including dataset type (real or simulated) and accessibility.

Study	Dataset	Availability
[9], [8]	Real dataset	DirecNet dataset [80], AI4PG dataset [81].
[16]	Real dataset collected from five T1D patients. Simulated dataset with 200 timestamps, 40 points per patient.	Not publicly available online. In-house dataset
[65]	Real dataset from two systems: 1) Guardian® Real-Time CGM System (Medtronic-Minimed) and 2) the FreeStyle Navigator® CGM System (Abbott Laboratories)	Not publicly available online. In-house dataset
[7]	Real dataset from eight T2D patients.	Not publicly available online. In-house dataset
[27]	Real dataset from thirteen T2D patients.	Not publicly available. In-house dataset
[66]	Real dataset from five T1D patients.	Not publicly available. In-house dataset
[58]	Simulation dataset	Not publicly available. The data is available upon request.
[72]	Real dataset available from Tidepool dataset. Simulated dataset	Tidepool available [83]
[12], [55]	Real dataset from 29 subjects, comprising 9 individuals with T1D and 20 healthy individuals	D1NAMO available [84]
[2]	Real dataset from 12 T1D patients	DirecNet dataset [80]
[57], [95], [24], [35], [14] [69], [21]	Real dataset from T1D patients	The dataset, OhioT1DM, is available upon reasonable request from Ohio University under a Data User Agreement [82].
[34], [33]	Real dataset from 27 T1D patients.	Not reported (in-house clinical data)
[63]	Read dataset of 10 T1D patients	Not reported (in-house clinical data)
[93]	Real dataset of 20 T1D patients	Not reported (in-house clinical data)

[94]	Simulated dataset virtual 9 T1D patient data generated by the UVA/Padova T1D simulator Real dataset of 20 T1D patients	Public (UVA/Padova simulator is open-source) Real dataset is not reported (in-house clinical data)
[47]	Real dataset of 24 T1D patients	Not reported (in-house clinical data)
[45]	Real dataset of 463 T1D patients	Not reported (in-house clinical data)
[31]	Real dataset of 11 T1D patients	Available in [90]
[17]	Simulated dataset virtual 9 T1D patient data generated by the UVA/Padova T1D simulator	Public (UVA/Padova simulator is open-source)
[68]	Real dataset of 124 T1D patients	Not reported (in-house clinical data)
[20]	Two real datasets: public: 6 T1D patients Private: 3 T2D patients	The private dataset: In-house dataset. The public real dataset: OhioT1DM, is available upon reasonable request from Ohio University under a Data User Agreement [82].
[6]	Simulated dataset virtual 30 T1D patient data generated by the UVA/Padova T1D simulator. Real dataset of 141 T1D patients	Public (UVA/Padova simulator is open-source) Real dataset is not reported (in-house clinical data)
[81]	Real dataset from 104 patients	Not reported (in-house clinical data)
[64], [15]	Real dataset from 10 patients	Not reported (in-house clinical data)
[51]	Real dataset from 9,584 hospital admissions for 6,187 patients	Not reported (in-house clinical data)
[61]	Two real datasets: First dataset: 124 T1D patients from Tidepool. Second dataset: 10 T1D patients	First dataset: Tidepool is available [83] Second dataset: Not reported (in-house clinical data)
[78]	Real dataset from 17,658 inpatients with diabetes, covering 32,758 hospital admissions.	Not reported (in-house clinical data)
[62]	Real dataset from 10 patients	Not reported (in-house clinical data)
[74]	Real dataset from 55 patients	Not reported (in-house clinical data)
[86]	Real dataset	Not reported (in-house clinical data)
[3], [5]	Real dataset from 768 T2D patients	Pima Indian Diabetes dataset [85], the hospital Frankfurt in Germany [86].
[29]	Simulated dataset generated using SimGlucose system, Real dataset collected from three volunteers using the DexCom SEVEN CGM system.	Not publicly available online. In-house dataset
[90], [71]	Real dataset from 768 T2D patients	Pima Indian Diabetes dataset [85]
[11], [60]	Real dataset from 300 T1D patient	Tidepool available [83]
[96]	Real dataset from 1 T1D patient Simulated dataset using the UVA/PADOVA T1D simulator	Simulated dataset is not publicly available. In-house dataset. The real dataset, OhioT1DM, is available upon reasonable request from Ohio University under a Data User Agreement [82].
[76]	Real dataset from five subjects of the T1D patients, T1DiabetesGranada.	Available online under specific permission at T1DiabetesGranada [89]
[97]	Real dataset of 12 T1D patients	Not reported (in-house clinical data)
[49]	Simulated dataset virtual 10 T1D patient data generated by the UVA/Padova T1D simulator. Two real datasets: First dataset: 10 T1D patients Second dataset: 6 T1D patients	Public (UVA/Padova simulator is open-source) First real dataset is not reported (in-house clinical data) Second real dataset is public: : OhioT1DM, is available upon reasonable request from Ohio University under a Data User Agreement [82].
[59]	Real dataset from 50 patients	Not reported (in-house clinical data)
[79]	Real dataset from 271 employees	Not publicly available. In-house dataset
[83]	Real dataset	Not publicly available. In-house dataset
[4]	Real dataset from 1 T1D patient	Available [88]
[60]	Real dataset available from Tidepool dataset.	Tidepool available [80]
[48]	Simulated dataset of 10 T1D patients generated by the UVA/Padova T1D simulator Read dataset set of 10 T1D patients	Not reported (in-house clinical data)
[23]	Real dataset from 112 patients	Not reported (in-house clinical data)
[80]	Two real datasets: First dataset: 34 T1D patients from DIAdvisor. Second dataset: 179 T1D patients from ChildrenData.	Not reported (in-house clinical data)